SERIE

APRENDER A

INVESTIGAR

Análisis de la Información



ADONAY MORENO GARZÓN

Instituto Colombiano para el Fomento de la Educación Superior
ICFES

Instituto Colombiano de Estudios Superiores de Incolda
ICESI

YOLANDA GALLARDO DE PARADA
ADONAY MORENO GARZÓN

Serie
APRENDER A INVESTIGAR

Módulo 4
ANÁLISIS DE LA INFORMACIÓN



INSTITUTO COLOMBIANO PARA EL FOMENTO
DE LA EDUCACIÓN SUPERIOR, ICFES

Subdirección General Técnica y de Fomento

PATRICIA MARTÍNEZ BARRIOS
Directora General

PATRICIA ASMAR AMADOR
Subdirectora General Técnica y de Fomento

MÓNICA IBARRA ROSERO
Jefe División de Fomento (A)

MARÍA JESÚS RESTREPO ALZATE
Coordinadora del Proyecto

Serie: APRENDER A INVESTIGAR
ISBN: 958-9279-11-2 Obra completa
ISBN: 958-9279-15-5 Módulo 4

1ª Edición: 1987
1ª Reimpresión: 1988
2ª Reimpresión: 1991
2ª Edición: 1995
Reimpresión: 1998
3ª Edición: (corregida y aumentada) 1999

Calle 17 Nº 3-40 A.A. 6319
Teléfono: 2819311 - 2834027 - 2834067 - 2435129
Fax: 2845309 - 2834047 - 2845980
Santa Fe de Bogotá

Diseño de carátula, diagramación e impresión:
ARFO EDITORES LTDA.
Carrera 15 Nº 53-86
Tels.: 2355968 - 2175794
Santa Fe de Bogotá, D.C.



INSTITUTO COLOMBIANO PARA EL FOMENTO DE LA EDUCACIÓN SUPERIOR

Serie: APRENDER A INVESTIGAR

Módulos:
1. CIENCIA, TECNOLOGÍA, SOCIEDAD Y DESARROLLO
2. LA INVESTIGACIÓN
3. RECOLECCIÓN DE LA INFORMACIÓN
4. ANÁLISIS DE LA INFORMACIÓN
5. EL PROYECTO DE INVESTIGACIÓN

La serie APRENDER A INVESTIGAR ha sido realizada por el ICFES. Para las ediciones anteriores se contó con el siguiente grupo de autores:

CARLOS ESCALANTE A.
Profesor Universidad Nacional de Colombia

HUMBERTO RODRÍGUEZ M.
Profesor Universidad Nacional de Colombia

ALBERTO MAYOR M.
Profesor Universidad Nacional de Colombia

EDUARDO VÉLEZ B.
Investigador Instituto SER de Investigaciones

ÁNGEL FACUNDO D.
Exjefe División de Fomento
Investigativo ICFES

El proyecto de actualización y revisión de la presente edición de la serie APRENDER A INVESTIGAR fue realizado por el ICFES, para lo cual se conformó el siguiente grupo de autores:

Módulo 1:
LUIS JAVIER JARAMILLO

Módulos 3 y 4:
ADONAY MORENO — Universidad San Buenaventura - Cali
YOLANDA GALLARDO DE PARADA — Universidad de Pamplona (N.S.)

Módulos 2 y 5:
MARIO TAMAYO Y TAMAYO — Universidad ICESI - Cali

Instructivos para videos:
LUZ ESTELLA URIBE VÉLEZ — EAFIT - Medellín

Contenido

SERIE APRENDER A INVESTIGAR
Presentación 7
Introducción 9
Propósito, población y objetivos de la serie 11
Estructura de aprendizaje de la serie 13
La organización de la serie: los módulos, y material audiovisual 17
Descripción sintética de los módulos 19
La asesoría de tutores 23

Módulo 4: ANÁLISIS DE LA INFORMACIÓN 25

1. NATURALEZA DE LA ESTADÍSTICA 25

2. DIVISIÓN DE LA ESTADÍSTICA 27
2.1 Estadística descriptiva 27
2.1.1 Descripción de datos, técnicas de representación gráfica 29
2.1.2 Distribución de frecuencias 31
2.1.3 Elaboración de una tabla de frecuencias 31
2.1.4 Presentación gráfica 33
2.1.4.1 Histogramas y polígonos de frecuencias 34
2.1.4.2 Gráficas de barras 35
2.1.4.3 Gráficas lineales 36
2.1.4.4 Gráficas circulares 37

3. DESCRIPCIÓN DE DATOS - TÉCNICAS NUMÉRICAS 41
3.1 Medidas de tendencia central 41
3.1.1 La moda 41
3.1.2 La mediana 42
3.1.3 La media aritmética 44
3.2 Medidas de dispersión 46
3.2.1 El rango 47
3.2.2 Percentiles 47
3.2.3 Varianza 48
3.2.4 Desviación estándar 50
3.2.5 Coeficiente de variación 50
3.3 Asimetría 51
3.4 Kurtosis 51
3.5 Tablas de contingencia 53

4. INTRODUCCIÓN A LAS PROBABILIDADES . . . 60
4.1 Probabilidades elementales . . . 60
4.2 Esperanza . . . 62
4.3 Leyes de las probabilidades . . . 62
4.4 Análisis combinatorio . . . 64
4.4.1 Permutaciones . . . 64
4.4.2 Variaciones . . . 66
4.4.3 Combinaciones . . . 66
4.5 Probabilidad condicional . . . 67
4.6 Distribución binomial . . . 70
4.7 Distribución normal . . . 70
4.8 Distribución de Poisson . . . 76

5. ESTADÍSTICA INFERENCIAL . . . 80
5.1 La prueba de hipótesis . . . 81
5.2 Pruebas de significancia de muestras únicas o simples . . . 86
5.2.1 Prueba Z . . . 86
5.2.2 Distribución t de Student . . . 87
5.2.3 Distribución Chi cuadrado: χ^2 . . . 92

6. REGRESIÓN Y CORRELACIÓN . . . 98
6.1 Introducción a la bidimensional . . . 98
6.2 Ajuste de una recta de regresión rectilínea simple . . . 101
6.3 Correlación . . . 104
6.4 Coeficiente de correlación . . . 107

7. ANÁLISIS DE LA VARIANZA . . . 111

8. ESTUDIO DE FACTIBILIDAD DEL SOFTWARE ESTADÍSTICO . . . 117
8.1 SAS . . . 117
8.2 SPSS . . . 119
8.3 Requerimientos mínimos para la instalación del Software SAS . . . 120
8.4 Requerimientos mínimos para la instalación del Software SPSS . . . 120
8.5 Manejador de Base de Datos . . . 121

9. STATGRAPHICS . . . 125
9.1 Grupo I: Data handling and system utilities . . . 125
9.2 Grupo II: Plotting and descriptive statistics . . . 125
9.3 Grupo III: Anova and regression analysis . . . 126
9.4 Grupo IV: Time series procedures . . . 126
9.5 Grupo V: Advanced procedures . . . 126
9.6 Grupo VI: Mathematical procedures . . . 126
9.7 Requerimientos físicos para la instalación de Statgraphics . . . 127

ANEXO: Instructivo para el uso del video
1. Uso didáctico del video . . . 137
2. Videos:
• La medición y las ciencias . . . 145
• La curva normal . . . 151
• La muestra . . . 157

BIBLIOGRAFÍA . . . 165

Módulo 4
Análisis de la información

El análisis de la información en el proceso investigativo, depende del enfoque y del tipo de investigación que se haya seleccionado, como también de los objetivos que se hayan planteado.

La estadística se constituye en una herramienta fundamental para el análisis de la información. Sin embargo es necesario precisar y seleccionar el tratamiento estadístico dependiendo del enfoque cuantitativo o cualitativo, de la escala de medición de las variables, de las hipótesis y de los objetivos.

La estadística es fundamental para resolver problemas de descripción de datos, análisis de muestras, contrastación de hipótesis, medición de relaciones y predicciones.

1. NATURALEZA DE LA ESTADÍSTICA

La estadística es una rama de la ciencia, encargada del diseño y aplicación de métodos para recolectar, organizar, analizar y hacer deducciones a partir de ellos.

Aunque los orígenes de la estadística se remontan a los estudios de los juegos de azar del siglo XVIII, sólo en los últimos 60 años se han desa-

rrollado las aplicaciones de los métodos estadísticos en casi todos los campos de la ciencia (social, comportamental y física). La mayor parte de las primeras aplicaciones de la estadística consistieron principalmente en la presentación de datos en forma de tablas y gráficas. Este campo se desarrolló con rapidez llegando a incluir gran variedad de métodos para ordenar, resumir y expresar en alguna forma las características de un conjunto de números. Hoy estas técnicas cubren lo que es sin duda la más visible aplicación de la estadística: la masa de información cuantitativa que se recopila y se publica todos los días en nuestra sociedad. Las tasas de natalidad, de mortalidad, los índices de precios, los promedios de goles obtenidos en las fechas semanales de fútbol, son algunas de las muchas «estadísticas» que nos resultan familiares.

Además de expresar las características de la información cuantificaba, estas medidas proporcionan una base importante para el análisis en casi todas las disciplinas académicas, especialmente en las ciencias sociales y del comportamiento, donde la conducta humana no puede generalmente ser descrita con la precisión que se consigue en las ciencias exactas. Las medidas estadísticas de satisfacción, inteligencia, aptitud para el trabajo y capacidad de liderazgo, que sirven para ampliar nuestro conocimiento de las motivaciones y del rendimiento humano, son ejemplos de lo dicho. Del mismo modo, los índices de precios, productividad, producto nacional bruto, empleo, reservas disponibles y exportaciones son elementos útiles tanto para el gerente como para el gobernante cuando se trata de trazar una política encaminada a lograr el desarrollo y la estabilidad económica a largo plazo.

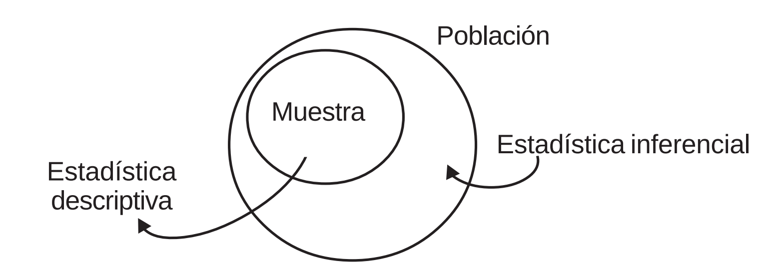
La estadística proporciona una metodología para evaluar y juzgar las discrepancias entre la realidad y la teoría. Además de su papel instrumental, el estudio de la estadística es importante para entender las posibilidades y limitaciones de la investigación experimental, para diferenciar las conclusiones que pueden obtenerse de los datos de aquellas que carecen de base empírica y en definitiva para desarrollar un pensamiento crítico y antidogmático ante la realidad.

En la actualidad con la ayuda de la informática y la tecnología el tratamiento estadístico de la información se hace más sencillo.

Para el análisis de datos cuantitativos, tenemos en la actualidad programas como el SAS, SPSS, MINITAB, SVSTAT, RESAMPLING, STATGRAPHIS. Para el análisis de datos cualitativos existen programas como el QUALPRO, ETHNOGRAPH, NUDIST, AQUAD.

2. DIVISIÓN DE LA ESTADÍSTICA

La estadística se divide en dos grandes ramas, dependiendo del método empleado para manejar la información y hacer que tenga sentido: estadística descriptiva y estadística inferencial.



Gráfica 1.

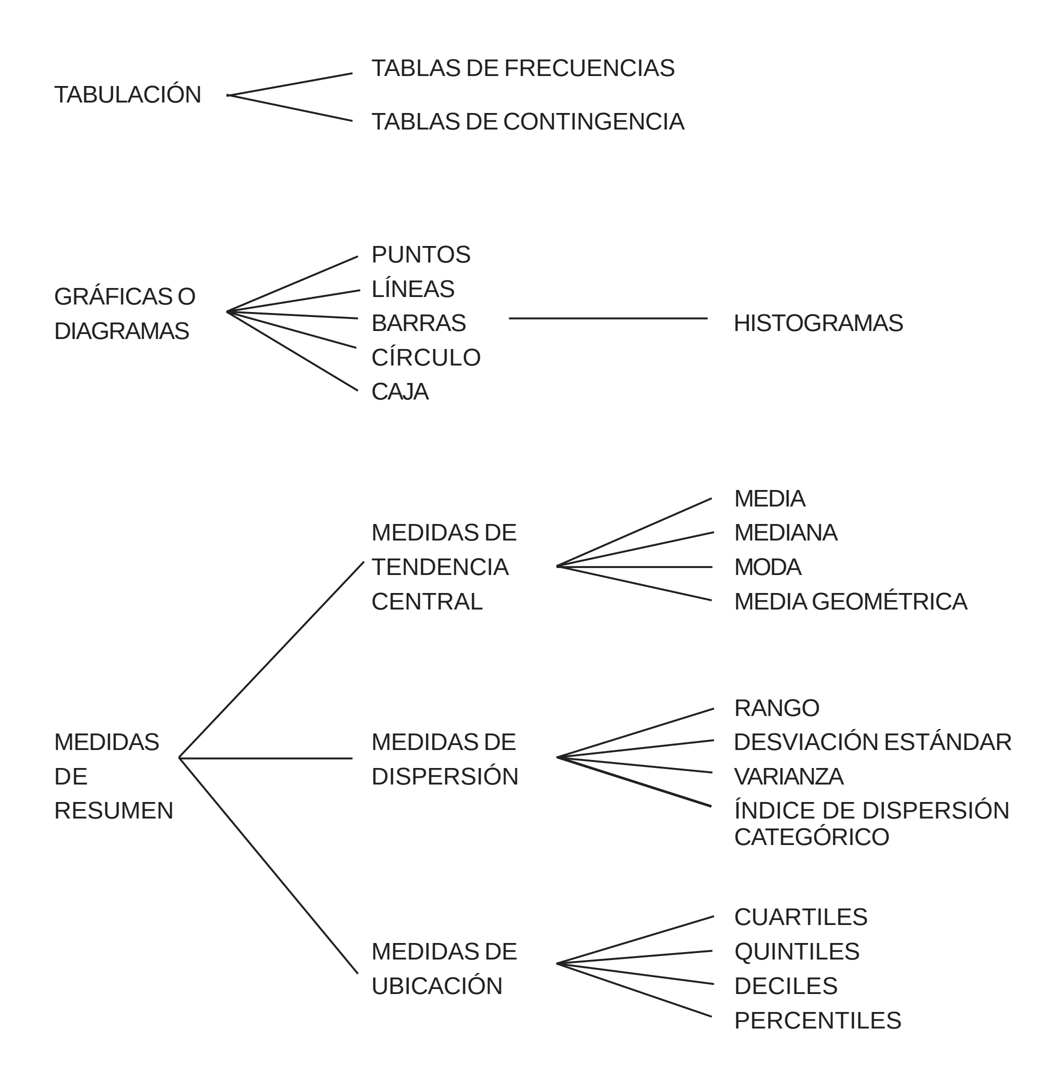
2.1 Estadística descriptiva. Permite describir resumir y analizar la información obtenida de la muestra.

Para tal fin se recolecta la información, se tabula, se grafica y en muchos casos en vez de trabajar con todas las observaciones, es preferible tener unas medidas que resuman los datos.

Básicamente hay tres tipos de medidas de resumen: medidas de tendencia central, medidas de dispersión o variabilidad de los datos y medidas de ubicación.

La gráfica 1 se interpreta así: con los datos de la muestra se aplica la estadística descriptiva; con los resultados de la estadística descriptiva se hacen análisis de estadística inferencial referidos a la población.

MAPA CONCEPTUAL DE ESTADÍSTICA DESCRIPTIVA



2.1.1 *Descripción de datos, técnicas de representación gráfica*

El concepto de la descripción está asociada a la distribución de frecuencias, que consiste en el ordenamiento o clasificación de los valores observados en una variable, de acuerdo con su magnitud numérica. Permite al investigador identificar la forma como ciertos puntos o características están distribuidos.

La distribución de frecuencia se puede construir a partir de variables medidas a cualquier nivel, desde nominal hasta de razón.

Podríamos suponer que queremos ver cuántos productores agrícolas en una vereda son grandes, medianos o pequeños productores.

Entrevistamos a los productores de la vereda y les preguntamos la extensión de su explotación y la clasificamos de acuerdo con el tamaño (magnitud) en una de las tres categorías.

Se construye la distribución de frecuencias que muestre cuántos productores son grandes, cuántos medianos y cuántos pequeños. El recuento será:

Tabla 1 - Tipo de productor agrícola

Tipo de productor	Frecuencia
Pequeño	□ ⊓ 7
Mediano	□ □ □ 12
Grande	⊠ □ 9
Total	28

En el cuadro sólo se ha agrupado el número de agricultores pertenecientes a cada categoría. El número de agricultores se denomina observaciones o frecuencia (f).

Para el caso de variables medidas a nivel intervalo o de razón, podemos hacer exactamente el mismo ejercicio.

Si se tiene, por ejemplo, la información acerca del número de cajas de madera que construyen al día unos carpinteros en un taller, podríamos clasificarlos por su productividad.

Tabla 2 - Número de cajas construidas por 15 carpinteros

Carpinteros	N° de cajas construidas
1	11
2	10
3	8
4	12
5	12
6	10
7	7
8	8
9	10
10	11
11	10
12	9
13	9
14	10
15	11

A partir de esta información podemos aclarar aún más la naturaleza de la producción por carpintero y se podría entrar a comparar esta productividad con la de otro tipo de taller o fábrica.

En la tabla 3 introducimos un mayor significado y ordenamiento a la información que se está analizando.

Tabla 3 - Número de cajas construidas por 15 carpinteros

Nº de cajas construidas	Nº de carpinteros que construyen X Nº de cajas	F
7	I	1
8	II	2
9	II	2
10	⧅	5
11	III	3
12	II	2

Hasta ahora los ordenamientos que se han hecho son bastante simples, pero en la vida real muchas veces la información analizada es más compleja y la distribución de frecuencias debe ser construida a partir de un mayor número de puntajes.

2.1.2 *Distribuciones de frecuencias*

La distribución de frecuencias es un método para organizar y resumir datos. Bajo este método los datos que componen una serie, se clasifican y ordenan, indicándose el número de veces en que se repite cada valor.

Los caracteres de los elementos de una población pueden ser: *cualitativos y cuantitativos.*

Los datos cualitativos, denominados también atributos, son todos aquellos fenómenos que pueden ser descritos cualitativamente, es decir mediante palabras; son ejemplos de atributos: la clasificación de los alumnos de una universidad por departamento de origen, clasificación de un grupo de personas por ocupación, por sexo, por cargo, etc.

Los caracteres cuantitativos, denominados variables, son todos aquellos fenómenos susceptibles de ser expresados cuantitativamente, es decir mediante números. Por ejemplo: peso, estatura, edad, número de hijos, salarios, etc.

Las variables se dividen en: discretas y continuas, teniendo en cuenta que ésta clasificación tiene más valor teórico que práctico.

Las variables discretas son aquellas que admiten solamente valores enteros, es decir no tienen valores intermedios; por ejemplo el número de hijos por familia, ya que no se puede decir que una familia tiene dos hijos y medio.

Las variables continuas, son aquellas que admiten valores fraccionarlos, pudiéndose establecer intervalos. Por ejemplo, la estatura de una persona que mide un metro con setenta centímetros, que pesa sesenta kilos, una libra y cuatro onzas, etc.

2.1.3 *Elaboración de una tabla de frecuencias*

Tomemos como ejemplo una muestra de 20 alumnos, determinando su peso en kilos; para facilitar el trabajo redondeamos las cifras.

$X_1 = 74$	$X_5 = 69$	$X_9 = 47$	$X_{13} = 65$	$X_{17} = 76$
$X_2 = 67$	$X_6 = 61$	$X_{10} = 85$	$X_{14} = 88$	$X_{18} = 57$
$X_3 = 94$	$X_7 = 71$	$X_{11} = 82$	$X_{15} = 52$	$X_{19} = 72$
$X_4 = 70$	$X_8 = 79$	$X_{12} = 55$	$X_{16} = 58$	$X_{20} = 66$

El primer paso a seguir consiste en determinar el valor máximo y el mínimo.

En el ejemplo tenemos:

$X_{máximo} = 94; X_{mínimo} = 47$

La diferencia entre el valor máximo y el mínimo se denomina recorrido o rango.

94 - 47 = 47

El rango, será entonces de 47.

Introducimos dos nuevos símbolos que son:

m = número de intervalos (o clase de la distribución)

c = amplitud de intervalo (o de clase de la distribución)

El valor de m, o sea, el número de intervalos se puede obtener mediante la siguiente fórmula:

m = 1 + 3.3 log n , donde n es el número de datos

Es de anotar que la anterior fórmula es poco conocida, por consiguiente, es poco usual. Sin embargo, se recomienda que el número de intervalos, hasta donde sea posible, no sea menor de 5 ni mayor de 16.

Para el cálculo de amplitud del intervalo se puede aplicar la siguiente fórmula:

$$C = \frac{X_{max.} - X_{min.}}{m} \quad \text{o sea} \quad C = \frac{\text{rango}}{m}$$

En el caso de nuestro ejemplo, el número de intervalos será:

m = 1 + 3.3 log2O de donde m = 1 + 3.3,x 1,30

m = 1 + 4,29

m = 5.29 ≅ 5 intervalos.

La amplitud del intervalo de clase será:

Rango R =47

Número de intervalos m = 5

$$C = \frac{R}{m} \quad \text{de donde} \quad C = \frac{47}{5} = 9.4$$

Redondeando y para mejor manejo de los datos, se puede considerar que la amplitud del intervalo sea de 9.

La tabla de frecuencias nos quedará en la siguiente forma,

Tabla 4 - Distribución de frecuencias del peso de 20 alumnos

Peso	Frecuencias (f)	Frecuencia Relativa Fr
46 - 55	3	3/20 = 0.15 = 15%
56 - 65	4	4/20 = 0.20 = 20%
66 - 75	7	7/20 = 0.35 = 35%
76 - 85	4	4/20 = 0.20 = 20%
86 - 95	2	2/20 = 0.10 = 10%
Total	**20**	**20/20 = 1 = 100%**

Frecuencia relativa (Fr)

Es su frecuencia dividida por el total de datos y se expresa generalmente como un porcentaje:

$$Fr = \frac{f}{n}$$

2.1.4 *Presentación gráfica*

Aunque las tablas sean ya el resultado de una concentración de datos, pueden ser, sin embargo, demasiado amplias y complejas, de modo que pierden una buena parte de lo que debería ser su cualidad primordial, la claridad.

Entonces, podemos recurrir a la presentación gráfica, para la mejor comprensión y análisis de los datos. En las variables discretas se hace la representación mediante diagramas de frecuencias; para ello, en el eje hori-

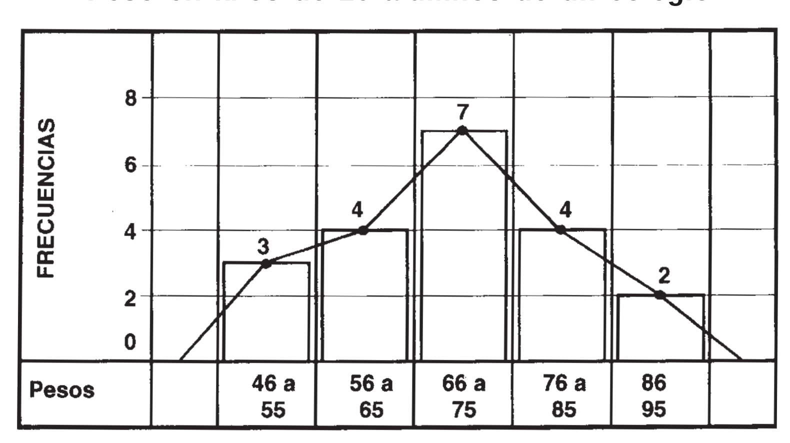
zontal, se colocan los distintos valores de las frecuencias absolutas o relativas. Si la representación se refiere a las frecuencias absolutas o relativas acumuladas se denomina: Diagrama de frecuencias acumuladas, colocándose los valores de la variable en el eje horizontal y las frecuencias Fr, en el vertical.

2.1.4.1 Histogramas y polígonos de frecuencias

Una distribución de frecuencias, se puede describir por medio de histogramas de frecuencias, que son gráficos que representan la información contenida en una distribución de frecuencias.

El ejemplo de la tabla 4, sobre el peso de los alumnos, se puede presentar en un histograma que representa exactamente lo mismo.

Gráfica 1.
Peso en kilos de 20 alumnos de un colegio



Otra manera de describir la distribución de la información obtenida, es por medio del polígono de frecuencias.

Estos son gráficos en la forma de una serie de líneas rectas conectadas entre sí y que unen puntos medios de intervalos a lo largo del eje horizontal.

El caso del peso de los alumnos, puede servirnos para ilustrar la técnica de los polígonos de frecuencia. Para construirlo se siguen los mismos pasos que para construir un histograma, pero en lugar de construir los rectángulos a partir de los límites superior e inferior de los intervalos, se calcula el punto medio del mismo y se unen, por medio de una línea recta que se conecta a los puntos medios de los demás intervalos.

Así por ejemplo en el gráfico 1 podemos trazar el polígono de frecuencias, uniendo los puntos medios del histograma de frecuencias.

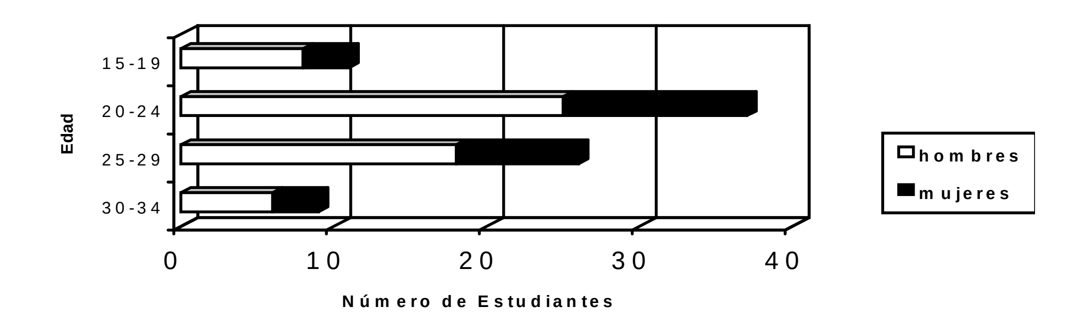
2.1.4.2 Gráficas de barras

La gráfica de barras es fácil de construir y su interpretación es de gran utilidad. Una gráfica diseñada para mostrar magnitudes absolutas deberá tener su inicio en cero y una escala de cantidades continuas. Las gráficas de barras pueden construirse en forma vertical u horizontal.

Las gráficas de barras pueden dividirse en tres tipos:

a. *Gráficas de barras con partes componentes.* En esta gráfica cada barra ha sido segmentada en dos partes componentes; así, por ejemplo, en la siguiente gráfica se tiene que la población de estudiantes de 100 semestre de la facultad X, de la Universidad Y, se distribuye según el sexo y la edad así:

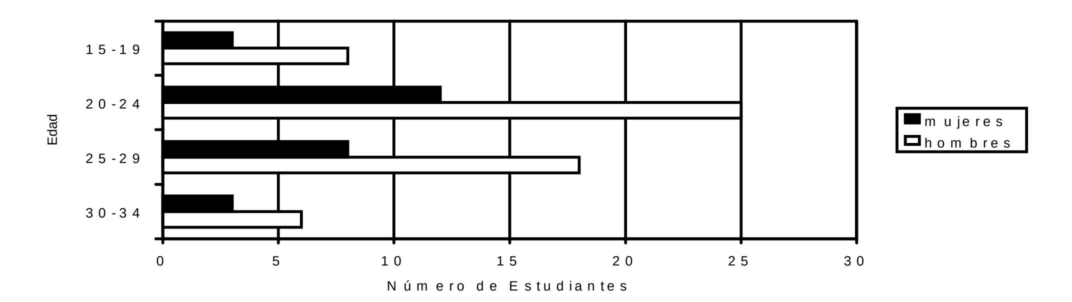
Distribución de los estudiantes de décimo semestre de la Facultad X, de la Universidad Y, por sexo y edad



En la gráfica se muestra el número de estudiantes clasificados por sexo y edades, en cada barra se presentan los dos componentes, los hombres y las mujeres.

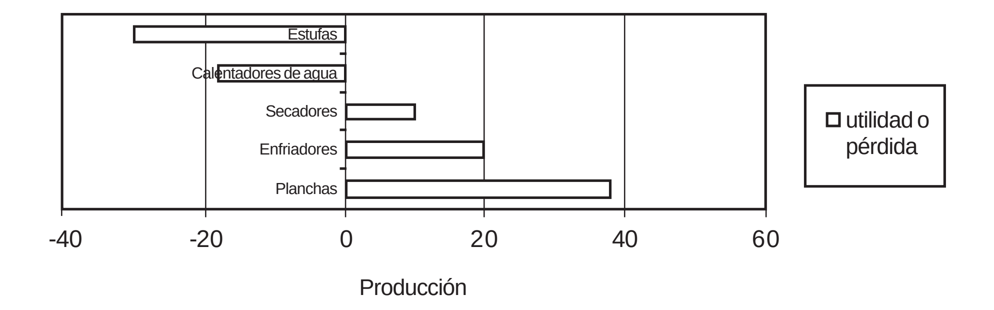
b. *Gráficas de barras agrupadas.* En esta gráfica, se presentan los mismos datos de estudiantes, excepto que los componentes por sexo se muestran separadamente así:

Distribución de los estudiantes de décimo semestre de la Facultad X, de la Universidad Y, según el sexo y la edad



c. *Gráficas de barras bidireccionales.* Cuando se desea graficar cantidades positivas y negativas, tales como pérdidas y ganancias en la producción de una empresa, por tipo de producto. Así, en la gráfica de producción de la empresa X, se tiene que las pérdidas y ganancias por tipo de producto son:

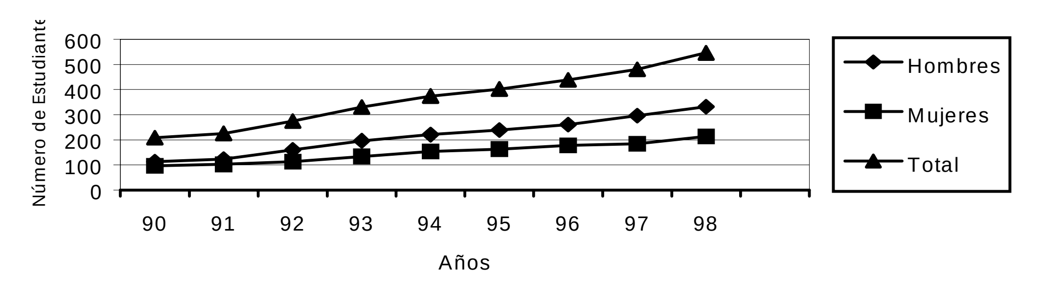
Utilidad o pérdida en la producción de la Empresa X



2.1.4.3 Gráficas lineales

Cuando debe presentarse una serie que cubre un gran número de períodos de tiempo, los datos graficados por medio de barras se ven demasiado acumulados; entonces, éstas pueden ser reemplazadas por una línea, ya que las líneas son más efectivas que las barras, cuando existen marcadas fluctuaciones en las series, o cuando deben presentarse varias series sobre la misma gráfica. Por ejemplo, el número de estudiantes de la facultad X, matriculados por sexo, en el período 1990-1998.

Número de estudiantes de la Facultad X, matriculados por sexo, en el período 1990-1998



2.1.4.4 Gráficas circulares

Las gráficas circulares, llamadas también de tipo pastel, se usan para mostrar los tamaños relativos de los componentes de un total. Son utilizados, para indicar cosas tales como participación de cada facultad en el total de estudiantes matriculados de determinada universidad, en períodos diferentes.

El proceso para realizar el diagrama consiste en una regla de tres.

Para conocer el ángulo de cada sector se debe relacionar los 3600 que tiene una circunferencia con el tamaño de la muestra y con cada una de sus frecuencias absolutas, así:

$$\begin{array}{l} 360° \longrightarrow n \\ X \longrightarrow F_1 \end{array}$$

Por ejemplo, si la siguiente tabla representa el número de docentes de una universidad, clasificados por modalidad educativa.

Modalidad	
Técnica	5
Tecnología	25
Universitaria	80
Post-grado	15
	N = 125

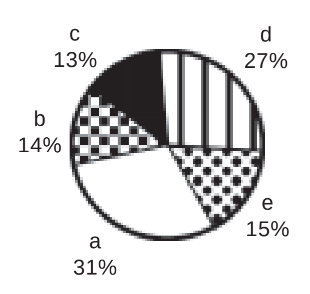
El diagrama circular se construirá así:

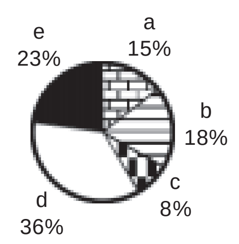
$$\begin{array}{ll} 360° \longrightarrow 125 \\ X \longrightarrow 5 \end{array}$$

Donde 125 es el tamaño de la muestra y 5 la frecuencia en la primera modalidad.

Luego $$X = \frac{360° \times 5}{125}$$

En el siguiente ejemplo, se muestra la distribución porcentual del número de estudiantes matriculados en 5 facultades de una universidad X, en los años de 1990 y 1998.





DESCRIPCIÓN DE DATOS, TÉCNICA DE REPRESENTACIÓN GRÁFICA - AUTOEVALUACIÓN N° 1

Lea cuidadosamente cada uno de los siguientes enunciados y señale la respuesta correcta.

1. El propósito principal de la estadística descriptiva es:
   a) Representar los datos por medio de histogramas.
   b) Obtener la información necesaria para la investigación.
   c) Medir en escalas nominales, ordinales e intervalos.
   d) Conocer las características generales de una distribución de frecuencias.
   e) Todas las anteriores.

2. Una distribución de frecuencias sólo se puede construir a partir de variables medidas en escalas intervalos.

   Sí _______ No _______

¿Por qué? ____

____

3. Si clasificamos 220 municipios en grandes, medianos y pequeños, de acuerdo con el número de habitantes, de forma tal que tenemos 49 grandes, 63 medianos y 108 pequeños, ¿cómo los representaría en un histograma de frecuencias como en un diagrama circular?

4. Para construir un histograma de frecuencias, primero es necesario construir un polígono de frecuencias.

   Sí _______ No _______

5. En un examen general de matemáticas los 30 alumnos de un curso obtuvieron las siguientes calificaciones:

Calificación	**N° de alumnos**
78	1
77	1
76	0
75	2
74	1
73	0
72	2
71	2
70	4
69	5
68	3
67	0
66	3
65	2
64	0
63	1
62	2
61	1

Represente la misma información en una distribución de frecuencias, basada en intervalos. Determine el rango, el número de intervalos y la amplitud de los intervalos. Represente los resultados por medio de un histograma y un polígono de frecuencias.

1. d

2. No

3. Gráfico

4. No

5. Rango = 17; número de intervalos = 6; amplitud de intervalo = 3

Calificaciones	F	Frecuencia Relativa %
61-63	4	13.3
64-66	5	16.7
67-69	8	26.7
70-72	8	26.7
73-75	3	10.0
76-78	30	100.0

Graficar

RESPUESTAS A LA AUTOEVALUACIÓN N° 1

3. DESCRIPCIÓN DE DATOS - TÉCNICAS NUMÉRICAS

La segunda técnica utilizada en la estadística descriptiva se basa en dos tipos de cálculos numéricos: Los de tendencia central y los de dispersión.

Las medidas de tendencia central se basan en que, en cualquier distribución de frecuencias, los valores tienden a concentrarse al rededor de un valor central de la distribución.

Las de dispersión, por el contrario, se basan precisamente en la manera en que los valores se distribuyen alrededor de esos valores centrales.

Para entender la naturaleza de la distribución, tenemos que tratar de describir el centro de la distribución de las mediciones y la forma como éstas varían alrededor de ese centro.

Podemos utilizar estas técnicas descriptivas, tanto para estudiar los parámetros de la población como las estadísticas de las muestras y en este último caso, a partir de ellas podemos estimar los correspondientes parámetros de la población.

3.1 Medidas de tendencia central

A continuación describimos las técnicas de tendencia central: moda, mediana y media aritmética.

3.1.1 *La moda*

La moda de una distribución se define como el valor que presenta la mayor frecuencia. Se usa con mediciones de escala nominal, ordinal, de intervalo o de razón.

Es comúnmente utilizada como una medida de popularidad, que refleja la tendencia de una opinión. En algunas distribuciones sólo hay una moda, pero en otras puede haber dos o más modas. Si tomamos 1, 4, 4, 4, 2, 5, 5, 8, 3, 6, 5, vemos que tanto el cuatro como el cinco aparecen con más frecuencia y en tres ocasiones. Es decir, hay dos modas y la distribución es bimodal.

Cuando se trabaja con datos agrupados, la moda se refiere como el valor medio del intervalo que constituye la mayor frecuencia.

Tabla 5. Puntajes obtenidos por 50 alumnos de un curso de física

Puntajes	Frecuencias
30 - 32	4
33 - 35	4
36 - 38	5
39 - 41	7
42 - 44	12
45 - 47	7
48 - 50	4
51 - 53	3
54 - 56	2
57 - 59	2
Total	**50**

En la tabla 5, presentamos la distribución obtenida al estudiar los puntajes obtenidos por 50 estudiantes como calificación en un curso de física.

El intervalo que contiene el mayor número de casos, o mayor frecuencia es 42 - 44. Este intervalo contiene los puntajes 42, 43 y 44. El valor medio del intervalo es, por lo tanto, 43 y lo denominamos como la moda.

3.1.2 *La mediana*

La mediana se define como la medida de tendencia central que divide a cualquier distribución en dos partes iguales.
En la siguiente distribución:

7, 8, 8, 9, 12, 15, 18, 18, 20, 21, 23.

La mediana es 15, porque se sitúa en el punto que divide la distribución en dos partes iguales. Hay el mismo número de casos antes y después del 15.

La mediana se usa en variables medidas en escala ordinal, intervalo o de razón. Su mayor uso es cuando se tienen muchas observaciones y generalmente se utiliza en distribuciones de ingresos, edades, pesos.

Cuando hay una distribución con un número par de casos, la mediana es el promedio de los dos valores medios. Así, en la siguiente distribución de notas:

78,95,86,73,52,90,89,84,76,92

El valor de la mediana es 85, porque, primero al ordenar la distribución de menor a mayor así:

52,73,76,78,84,86,89,90,92,95

Siendo 10 el total de notas, las que aparecen en la posición quinta y sexta están en la mitad de la distribución, entonces la mediana será:

$$\frac{84 + 86}{2} = 85$$

Cuando se tiene información agrupada, la mediana se define como el valor dentro del intervalo que divide la distribución en dos partes iguales. El símbolo utilizado es Me.

En la tabla 6, tenemos una distribución cualquiera con 5 intervalos de tamaño 5 y con 30 observaciones.

Tabla 6. Distribución de frecuencias de los pesos de 30 estudiantes

Intervalos	F	Frecuencias acumuladas
30 - 34	4	4
35 - 39	7	11
40 - 44	8	19
45 - 49	6	25
50 - 54	5	30
Total	**30**	

Los pasos a seguir para el cálculo de la mediana son:

- Encuentro las frecuencias absolutas acumuladas.
- Con base en las frecuencias acumuladas ubico el intervalo donde quede la frecuencias correspondiente a la mitad del tamaño de la muestra, es decir: $\frac{n}{2}$
- Encuentro el valor del límite real inferior del intervalo donde está $\frac{n}{2}$
- Aplico la siguiente fórmula:

$$\text{Mediana} = Li + \frac{\frac{n}{2} - \Sigma Fa}{F\frac{n}{2}} \, C$$

donde Li = Límite real inferior donde está ubicada $\frac{n}{2}$

ΣFa = Suma de las frecuencias anteriores a donde está ubicado $\frac{n}{2}$

C = Amplitud del intervalo.

En el ejemplo anterior de los pesos de 30 estudiantes la mediana está dada por:

$$\frac{n}{2} = \frac{30}{2} = 15$$

Li = 39.5 Corresponde al límite inferior donde está 15 como frecuencia acumulada.

$F\frac{n}{2} = 8$ Corresponde a la frecuencia del intervalo donde está $\frac{n}{2}$

$\Sigma Fa = 11$ Corresponde a la suma de las frecuencias antes de $\frac{n}{2}$

C = 5 Es la amplitud del intervalo.

Luego:

$$\text{Mediana} = 39.5 + \frac{(15 - 11)}{8} \times 5$$

Mediana = 42

3.1.3 *La media aritmética*

Es la medida de tendencia central más conocida, es fácil de calcular, de gran estabilidad en el muestreo, sus fórmulas permiten tratamiento algebraico.

Además, presenta el inconveniente de ser muy sensible a los valores extremos, cuando éstos son demasiado bajos o demasiado altos. Se representa así: $\overline{X}$

La media aritmética se define como la suma de todos los valores observados dividido por el número de observaciones (n).

La fórmula para datos no agrupados es: $\overline{X} = \frac{\Sigma xi}{n}$

Donde **Σ**-Xi corresponde a la sumatoria de todos los valores de la muestra.

La media aritmética de la siguiente distribución es:

1.1, 2, 1, 2.4, 3, 2, 4.5

$$\overline{X} = \frac{1.1 + 2 + 1 + 2.4 + 3 + 2 + 4.5}{7} = \frac{16.1}{7} = 2.28$$

Para las distribuciones con datos agregados, existe una fórmula, aunque un poco más complicada, es bastante fácil de aplicar.

$$\text{media} = \frac{\Sigma fx}{n}$$

donde f corresponde a las frecuencias, y X al valor de cada marca de clase.

Tabla 7. De valores de los pesos de los 30 estudiantes

Intervalos	x Marca de clase	f Frecuencias	fx
30-34	32	4	128
35-39	37	7	259
40-44	42	8	336
45-49	47	6	282
50-54	52	5	260
	Total	**30**	**1265**

La marca de clase es el punto medio de cada intervalo. Para tal fin se suman los dos valores extremos y se divide entre dos. Por ejemplo, la marea de clase para el primer intervalo es:

$$X = \frac{30 + 34}{2} = \frac{64}{2} = 32$$

Por consiguiente al aplicar la fórmula la media aritmética es:

$$media = \frac{1.265}{30} = 42.17$$

La media aritmética tiene sus ventajas, sobre las otras medidas de tendencia central, en que:

- Cada caso se incluye en el cálculo.
- Es rígidamente calculada, es decir, sólo hay una para conjunto de datos.
- Es importante para hacer inferencias, es decir sirve también para calcular parámetros de la población.

La media aritmética sólo se puede calcular a valores numéricos, es decir que estén en escala de intervalos o de razón.

3.2 Medidas de dispersión

Son aquellas que nos determinan cómo se agrupan o se dispersan los datos alrededor de un promedio.

En este capítulo se estudiarán las medidas que se utilizan para determinar cuán bien representan los promedios a la distribución considerada.

En el siguiente ejemplo se presentan las notas obtenidas por dos grupos de estudiantes:

Grupo A	Grupo B
80	97
80	95
75	70
83	72
82	73
81	96
82	80
75	72
79	71
7	71
795	797

$\overline{X}_A = 79.5$
$\overline{X}_B = 79.7$

La media aritmética de ambas series es prácticamente igual (79.5 en el grupo A y 79.7 en el grupo B). Un análisis de las cifras individuales revelan un gran contraste. En el grupo A hubo muy poca variación entre las notas, siendo la más alta 83 y la más baja 75. En el grupo B, se nota mayor variación, en este grupo la mayor nota fue 97 y la menor 71. Como conclusión se podría decir que en el grupo B hubo notas muy altas y muy bajas. En el grupo A las notas tuvieron una mayor concentración alrededor del promedio.

Para establecer esta característica se utilizan las medidas de variación o dispersión, entre las cuales las más utilizadas son el rango, la varianza y la desviación estándar.

3.2.1 *El rango*

La medida más simple de dispersión es el rango. Este identifica la distancia entre el valor mayor y el valor menor de la distribución. Más específicamente, se define como la diferencia entre el mayor valor y el menor valor. Se simboliza por **r**.

Ejemplo, el rango de la siguiente distribución es:

25, 36, 64, 20, 48, 59.

r = 64 - 20 = 44

El rango es sencillo de calcular pero tiene la desventaja de que es sensible a los valores extremos.

3.2.2 *Percentiles*

Los percentiles son usados para calcular una segunda medida de dispersión. El P-ésimo percentil de un conjunto de mediciones ordenadas según su magnitud, es el valor que tiene P% de las mediciones por debajo de él y (100 P%) por encima.

Se utilizan muy frecuentemente para describir los resultados de pruebas de conocimiento, como los del Sistema Nacional de Pruebas y en la clasificación de personas en forma comparativa. Entre los percentiles de más interés están el 25, el 50 y el 75, frecuentemente denominados como el menor cuartil Q_1, el cuartil mediano (mediana) Q_2 y el cuartil superior, Q_3.

3.2.3 *Varianza*

De todas las medidas de dispersión, la más importante, más conocida y usada es la varianza. Se define como la media aritmética de los cuadrados de las desviaciones, respecto a su media. Se simboliza por S^2.

Su fórmula es:

$$S^2 = \frac{\sum (x_i - \text{media})^2}{n}$$ para datos no agrupados

$$S^2 = \frac{\sum f\,(X - \text{media})^2}{n}$$ datos agrupados

Ejemplo con datos no agrupados

$X_i = 5;\ X_2 = 3;\ X_3 = 1;\ X_4 = 6;\ X_5 = 10$

$$\overline{X} = \frac{5 + 3 + 1 + 6 + 10}{5} = \frac{25}{5} = 5$$

$$S^2 = \frac{(5\text{-}5)^2 + (3\text{-}5)^2 + (1\text{-}5)^2 + (6\text{-}5)^2 + (10\text{-}5)^2}{5}$$

$$S^2 = \frac{0 + 4 + 16 + 1 + 25}{5} = \frac{46}{5} = 9.2$$

Lo cual indica que la varianza de los datos es de 9.2

Ejemplo para datos agrupados

Supongamos la siguiente tabla de pesos en kilogramos:

Tabla 8 - Pesos de 100 estudiantes de un colegio

Pesos	**Marca de clase X**	**$X - \bar{X}$**	**$(X - \bar{X})^2$**	**Frecuencias f**	**$f(X - \bar{X})^2$**
60-62	61	-6.45	41.60	5	208.0
63-65	64	-3.45	11.90	18	214.2
66-68	67	-0.45	0.20	42	8.4
69-71	70	2.55	6.50	27	175.5
72-74	73	5.55	30.80	8	246.4
Total				**100 = n**	**852.51**

$\bar{X} = 67.45$

$$S^2 = \frac{\sum f(\bar{x} - x)^2}{n}$$

donde X es la marca de clase
$\bar{X}$ la media aritmética
f la frecuencia de cada intervalo

$$S^2 = \frac{852.\bar{5}}{100} = 8.52$$

Propiedades de la varianza

a. La varianza siempre debe ser un valor positivo $S^2 > 0$

b. La varianza de una constante es 0, es decir: si Xi = C para todo i, entonces $S^2 = 0$

c. La varianza de una constante más una variable, es igual a la varianza de la variable.

$S^2_{(k + X)}: = S^2_k + S^2_x = 0 + S^2_x = S^2_x$

También es válida para la diferencia

$S^2_{(k + X)}: = S^2_k + S^2_x = 0 - S^2_x = S^2_x$

d. La varianza de una constante por una variable, es igual al producto de la constante al cuadrado por la varianza de la variable:

$S^2_{(kX)}: = k^2 S^2_X$

3.2.4 *Desviación estándar*

La desviación típica estándar es la raíz cuadrada de la varianza, considerada siempre con signo positivo.

$$S = \sqrt{S^2}$$

Para el caso del peso de los 100 estudiantes del colegio la desviación estándar es:

$$S = \sqrt{8.52} = 2.92$$

La varianza se expresa siempre en unidades diferentes a las originales, es decir, si la variable se refiere a peso en Kg, al calcular la varianza estará dado el peso en Kg al cuadrado. Es una de las razones por la cual se utiliza la desviación estándar, pues se expresa en las mismas unidades de la variable.

3.2.5 *Coeficiente de variación*

Esta medida relaciona la desviación estándar y la media, para expresar la variación de la desviación con respecto a la media aritmética. Se acostumbra expresarlo en porcentaje.

La fórmula que se utiliza es:

$$Cv = \frac{S}{\overline{X}} \bullet 100$$

El coeficiente de variación sirve para determinar el grado de homogeneidad de la información. Si el valor del coeficiente de variación es pequeño indica que la información tiene un alto grado de homogeneidad y si el coeficiente de variación es grande es porque la información es heterogénea.

Ejemplo: al hallar el coeficiente de variación de 6, 3, 4, 7, 8

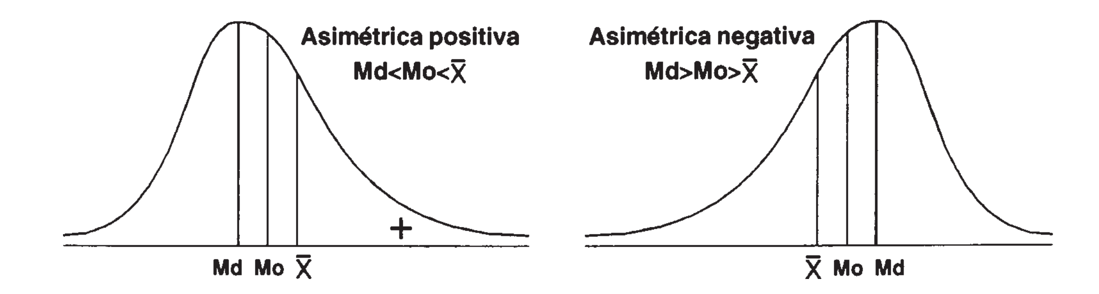
$$\overline{X} = \frac{28}{5} = 5.6$$

$$S = 1.85 \qquad Cv = \frac{1.85}{5.6} = 0.3304 = 33.04\%$$

Lo cual indica que la información es homogénea, pues el coeficiente de variación es de 33.04%.

3.3 Asimetría

Una distribución es simétrica si se tiene que: la media es igual diana e igual al modo ($\bar{X}$ = Md = Mo); pero si la distribución se vuelve asimétrica las tres medidas se separan y entonces el valor promedio ar será mayor que la mediana, que a su vez será mayor que el modo, deduciéndose que la distribución es asimétrica positiva. Si la media aritmética es menor que la mediana y ésta menor que el modo, se dice que la distribución es asimétrica negativa. En la distribución asimétrica positiva, la curva presenta un alargamiento a la derecha; en la negativa el alargamiento presenta hacia la izquierda. Véase gráficas 4.



Gráfica 4. Asimetría

Las fórmulas para calcular la asimetría y su grado son:

$$As = \frac{\text{Media aritmética - Modo}}{\text{Desviación estándar}} = \frac{\bar{X} - Mo}{s}$$

$$As = \frac{\text{3(Media aritmética - Mediana)}}{\text{Desviación estándar}} = \frac{3(\bar{X} - Md)}{s}$$

3.4 Kurtosis

Una característica importante de la variación de algunas distribuciones es su grado de agudeza en la cima de la curva que las representa agudeza que se observa en la región del modo, comparada con las condiciones halladas para el mismo sitio en la curva normal, es lo que Kurtosis.

Si la curva es más plana que la normal, la distribución se llama achatada o platicúrtica; si es más aguda, lleva el nombre de apuntada o leptocúrtica, y si la curva es normal, se denomina mesocúrtica.

La Kurtosis es una medida de altura de la curva, por lo tanto está representada por el cuarto momento de la media. En la misma forma que para la asimetría, su cálculo se efectúa en función de la desviación estándar y de los momentos unidimensionales de orden cuatro con respecto a la media aritmética.

Su fórmula es: $a_4 = \dfrac{m_4}{s^4} = \dfrac{m_4}{m^2_2}$

Otra formula para medir la Kurtosis es la que se basa en los cuartiles y percentiles.

$$K = \frac{Q}{P_{90} - P_{10}}$$

En donde $Q = 1/2\ (Q_3 - Q_1)$. Si el valor encontrado de la Kurtosis es igual a 0.263 se dice que la curva es mesocúrtica o normal; si es mayor, la curva es leptocúrtica y si es menor, la curva es platicúrtica.

Por ejemplo, calcular el coeficiente de Kurtosis, si el valor del primer cuartil es $Q_1 = 68.25$; $Q_3 = 90.75$; $P_{10}\ D_1 = 58.12$ y $P_{90} = D_9 = 101.00$

Aplicando la fórmula se tiene:

$$K = \frac{Q}{P_{90} - P_{10}} \quad \text{en donde } Q = \frac{1}{2}(Q_3 - Q_1) \quad \text{reemplazando}$$

$$\text{Se tiene } Q = \frac{1}{2}\ (90.75 - 68.25) = \frac{1}{2}\ (22.5) = 11.25$$

$$K = \frac{11.25}{101.00 - 58.12} = \frac{11.25}{42.88} = 0.262$$

Se puede decir que la curva es normal o mesocúrtica.

3.5 Tablas de contingencia

En muchas ocasiones, la información puede organizarse y presentarse de acuerdo con la relación entre dos o más variables.

Es cuando se habla de tablas de doble entrada o múltiples entradas o comparación de colectivos. Las variables pueden ser categóricas o numéricas.

Esta forma de representación presenta grandes ventajas, entre otras resume información y posibilita el análisis de dependencia, asociación, correlación y homogeneidad.

Supongamos que a partir de datos disponibles en nuestro estudio se ha elaborado el siguiente cuadro:

Tabla 9. Distribución de alumnos por jornadas y clase de colegio

CLASIFICACIÓN DE LA JORNADA	NÚMERO DE ALUMNOS		
	COLEGIOS OFICIALES	COLEGIOS PRIVADOS	TOTAL
Única	230	840	1070
Mañana	583	614	1197
Tarde	418	234	652
Total	**1231**	**1688**	**2919**

Si comparamos los datos en dirección horizontal constatamos que el mayor número de estudiantes corresponde a la jornada de la mañana, mientras que si leemos los datos en dirección vertical el mayor número de estudiantes corresponde los colegios privados.

En ocasiones el investigador desea especificar y analizar los datos con más variables. Para el ejemplo anterior podría ser sexo y edad.

Tabla 10. Distribución de alumnos por jornadas, clase de colegio, sexo y edad

Clasificación de la jornada	COLEGIOS OFICIALES				COLEGIOS PRIVADOS			
	Hombres		Mujeres		Hombres		Mujeres	
	10-15	**15-20**	**10-15**	**15-20**	**10-15**	**15-20**	**10-15**	**15-20**
Única	52	80	22	76	132	180	423	105
Mañana	121	47	168	247	241	132	86	155
Tarde	123	154	49	92	76	51	47	60
Subtotal	**296**	**281**	**239**	**415**	**449**	**363**	**5561**	**320**
TOTAL	**1231**				**1688**			

En este ejemplo, el 52 corresponde al número de estudiantes hombres entre los 10 y 15 años, de jornada única pertenecientes a colegios oficiales.

Las tablas de contingencia se pueden realizar con frecuencias relativas, es decir, con porcentajes.

ÍNDICE DE DISPERSIÓN CATEGÓRICO

Para datos ordinales y nominales se utiliza el llamado ÍNDICE DE DISPERSIÓN CATEGÓRICO, cuya fórmula es:

$$D = \frac{K(N^2 - \Sigma f^2)}{N^2(K - 1)}$$

Donde K es el número de categorías de la variable.
Σf^2 es la suma de las frecuencias al cuadrado
N = número de casos (tamaño de la muestra)

Por ejemplo:

Si tenemos la siguiente clasificación de 15 estudiantes por preferencias deportivas:

DEPORTES	ALUMNOS	f^2
FÚTBOL	8	64
VOLEIBOL	2	4
BASKET	5	25
TOTAL	15	

El índice de dispersión categórico sería:

$$D = \frac{3(225 - 93)}{225(3 - 1)}$$

$$D = \frac{396}{450} = 0.88$$

Lo cual indica que los datos presentan una dispersión del 88%.

DISTRIBUCIÓN DE DATOS, TÉCNICAS NUMÉRICAS
AUTOEVALUACIÓN N° 2

Lea cuidadosamente cada uno de los siguientes enunciados y señale la respuesta correcta.

1. Para entender la naturaleza de la distribución de frecuencias, las medidas de tendencia central son mejores que las de dispersión

   Sí ________ No ________

2. Las medidas de tendencia central son-.

   ____

3. ¿Cuál es la mediana de la siguiente distribución?

   4, 4, 5, 3, 7, 9, 5, 1, 7, 3, 7, 7, 5

4. Un profesor pidió los pesos en Kg a 30 estudiantes del curso de álgebra y obtuvo los siguientes resultados:

Edad	**No. de estudiantes f**
30-34	5
35-39	6
40-44	8
45-49	7
50-54	4
Total	**30**

a. ¿Cuál es el peso promedio de sus alumnos? __________

b. ¿Cuál es la mediana? __________

c. ¿Cuál es la moda de las edades? __________

5. ¿Cuáles son las modas de las siguientes series de números?

a) 122, 103, 148, 113, 103, 118.

b) 75, 84, 37, 49, 51, 83.

c) 22, 43, 22, 38, 41, 43, 20.

6. ¿Cuál es la medida de dispersión más sencilla?

____

7. El rango tiene la desventaja de ser sensible a los valores extremos.

Sí ________ No ________

8. La varianza siempre debe ser un valor positivo.

   Sí _______ No _______

9. ¿Cómo se define la Desviación Estándar?

   ____

10. Una de las características de la Kurtosis es el grado de agudeza de la curva, la cual se puede clasificar en tres tipos:

    a. Una más plana que la normal 0: ____________________
    b. La segunda más aguda 0: ____________________
    c. Si la curva es normal 0: ____________________

RESPUESTA A LA AUTOEVALUACIÓN N° 2

1. No
2. Media, mediana y moda
3. 5
4. a. $\overline{X}$ = 4.18
   b. Mediana = 42
   c. moda= 8
5. a) 103 b) No hay c) 22 y 43
6. El rango
7. Sí
8. Sí
9. La raíz cuadrada de la varianza
10. a. Platicúrtica. b. Leptocúrtica. c. Mesocúrtica

Taller de estadística N° 1

Se realizó un experimento para evaluar el efecto de la edad en la frecuencia cardíaca, cuando se somete una persona a un grado específico de ejercicios. Se seleccionaron al azar hombres de cuatro grupos de edades, 10-19, 20-39, 40-59, 60-69. Cada individuo recorrió la banda sin fin a una velocidad específica durante 12 minutos y se registró el aumento de la frecuencia cardíaca, la diferencia antes y después del ejercicio, (en latidos por minuto)

¿Qué puede usted concluir, si los siguientes datos son el resultado del procesamiento de la información apoyada en el computador.

VARIABLE	GRUPO 1	GRUPO 2	GRUPO 3	GRUPO 4
Talla	15	10	12	12
Promedio	30.1333	27.5	29.5	38.3333
Mediana	30	27.5	29	37.5
Moda	22	24	33	36
Varianza	24.8381	23.8333	19.1818	50.6061
Desviación estándar	4.98378	4.88194	4.37971	7.11379
Mínimo	22	21	22	28
Máximo	39	34	37	54
Rango	17	13	1.5	26
Quartil inferior	26	24	26.5	34.5
Quartil superior	34	32	33	42
Rango interquartil	8	8	6.5	7.5
Skewness	-0.0873053	-0.0716212	0.0311649	0.668752
Kurtosis	-0.612727	-1.59948	-0.726827	1.02909
Coeficiente de variación	16.5391	17.7525	14.8465	18.5577
Suma	452	275	354	460

Preguntas

1. ¿Qué tipo de muestreo se realizó?
2. ¿Qué grupo fue más homogéneo en los resultados? ¿Por qué?
3. Comparando las medias aritméticas ¿qué grupo obtuvo la menor media?
4. ¿Qué grupo presenta un mayor rango? ¿Qué significa?
5. Si consideramos el grupo 4
   5.1 ¿Cuál es el mínimo valor?
   5.2 ¿Cuál fue el mayor aumento de frecuencia cardíaca?
   5.3 ¿Qué porcentaje de la información está por debajo de 34.5?

5.4 ¿Qué porcentaje de la información está por encima de 34.5?
5.5 ¿Qué porcentaje de la información está entre 34.5 y 42?

6. Si consideramos el grupo 2
   6.1 ¿Qué porcentaje de la información está por encima de 32?
   6.2 ¿Qué valor corresponde al dato que más se repite?
   6.3 ¿Qué dato está ubicado en el primer cuartil?
   6.4 ¿Qué significado tienen las dos respuestas anteriores?

Respuestas

1. Muestreo aleatorio por estratos según edades.
2. El grupo No. 3 porque obtuvo el menor coeficiente de variación.
3. El grupo No. 2 con un valor de 27.5.
4. El grupo No. 4 significa que obtuvo mayor variabilidad entre el dato mayor y el dato menor.
5. 5.1 28
   5.2 54
   5.3 El 25%
   5.4 El 75%
   5.5 El 50%
6. 6.1 El 25%
   6.2 24
   6.3 24
   6.4 Significa que por debajo del dato que más se repite hay un 25% de la información.

4. INTRODUCCIÓN A LAS PROBABILIDADES

4.1 Probabilidades elementales

Consideramos varios ejemplos: es casi seguro que hoy no llueve, con ello expresamos algún grado de confianza de que no llueve, a pesar de la completa seguridad.

¿Qué posibilidad existe de obtener una buena nota en el examen de economía? ¿En el de estadística?

«En el próximo examen tengo la oportunidad de obtener una nota alta».

Si un aficionado a la hípica es preguntado sobre una fija en una determinada carrera, puede darnos una insinuación de cuál caballo puede ganar. «El caballo X tiene una buena oportunidad».

Nos encontramos con un conjunto cuya técnica se refiere a resultados que no estamos muy seguros. Asociamos cierto grado de confianza, con relación a que cada una de estas predicciones se realice. El cierto grado de confianza reposa en la experiencia (pasada y presente).

Los Bayecianos definen la probabilidad de un acontecimiento, como una medida del grado de confianza que uno tiene, en que ocurra el acontecimiento.

Otros autores consideran que la probabilidad pertenece a aquellas nociones imposibles de ser definidas adecuadamente y se les considera basadas en la experiencia. Por lo tanto, se puede decir que la probabilidad es una creencia basada en la experiencia.

Consideramos algunas definiciones, según el método.

a. **Método axiomático.** El cual concibe la probabilidad de ocurrencia de un suceso, como un número comprendido entre cero y uno. Este concepto tiene que ver directamente con la noción de frecuencia relativa, donde: $0 < p < l$.

Supongamos que se lanza 100 veces una moneda, anotamos el número de veces que sale cara y las veces que sale sello; los resultados fueron los siguientes:

Frecuencias absolutas:	cara 56 veces;	sello 44 veces
Frecuencias relativas:	56/100;	44/100
Posibilidad:	p = 56% (éxito);	q = 44% (fracaso)

b. **Método empírico o práctico.** Considera la probabilidad de un suceso, como aquel número al cual se aproxima cada vez la frecuencia relativa de la ocurrencia de un suceso, cuando las veces que se repite el experimento que origina ese suceso es bastante grande. Este concepto tiene algo que ver con el experimento de Quetelet, en donde la probabilidad de un suceso tiende a estabilizarse en un punto, cuando el número de experimentos se va haciendo cada vez más grande.

c. **Método clásico.** Considera la probabilidad como el cuociente de dividir los casos favorables, que pueden ocurrir en un suceso, por el total de casos posibles.

$$p = \frac{\text{Número de casos favorables}}{\text{Número de casos posibles}}$$

Ejemplos:

– En el lanzamiento de una moneda hay:

2 posibilidades	**N° Combinaciones**	**Probabilidades**
C	(C)	1/2 = 0.5
S	(S)	1/2 = 0.5
	2	1 = 1.0

– Lanzamiento de dos monedas.

4 Posibilidades	**N° Combinaciones**		**Probabilidades**
2C	(C,C)	1	1/4 = 0.25
1 C y 1 S	(C, S)	(S, C) 2	2/4 = 0.50
2S	(S,S)	1	1/4 = 0.25
		4	1 = 1.00

– Lanzamiento de tres monedas.

8 posibilidades	**N° Combinaciones**		**Probabilidades**
1C	(C, C, C)	1	1/8 = 0.125
2C y 1S	(C, C, S) (C, S, C) (S, C, C)	3	3/8 = 0.375
2S y 1C	(S,S,C)(S,C,S)(C,S,S)	3	3/8 = 0.375
3S	(S, S, S)	1	1/8 = 0.125
		8	1 = 1.000

En efecto el número de casos posibles en los ejemplos anteriores se obtiene:

$2^1 = 1$ 2^n
$2^2 = 4$ 2 = casos posibles en una moneda
$2^3 = 8$ n = número de lanzamientos

En un dado sería: 6^n

$6^1 = 6$ 6 = casos posibles en un dado
$6^2 = 36$ 36 = casos posibles al lanzar dos dados

4.2 Esperanza

Si p es la posibilidad de éxito en un suceso, en un solo ensayo, el número de sucesos o esperanzas de ese suceso en n ensayos, estará dado por el producto de n y la probabilidad de éxito p.

$$E = n\,p$$

Ejemplo:

En el lanzamiento 900 veces de dos dados, ¿cuál es la esperanza de que la suma de sus caras sea un valor menor a 6?

Primero obtenemos la probabilidad de éxito del suceso en un solo ensayo.

(1,1) (1,2) (1,3) (1,4) (2,1) (2,2) (2,3) (3,1) (3,2) (4,1) = 10

Como se lanzan 900 veces, en dos dados, se tiene:

$$E = np = 900\left\{\frac{10}{36}\right\} = \frac{900}{36} = 250$$

La esperanza es que en 250 de los 900 lanzamientos, la suma de sus caras sea menor de 6.

4.3 Leyes de las probabilidades

Para facilitar el cálculo de las probabilidades se emplean estas leyes.

1. Si dos o más sucesos son tales que solamente uno de ellos puede ocurrir en un solo ensayo, se dice que son mutuamente excluyentes. Se

denomina probabilidad aditiva y será igual a la suma de las probabilidades de cada suceso.

$P = P_1 + P_2 + P_3 .... + P_n$

Ejemplo:

La probabilidad de obtener un as o un rey, sacando una sola carta en una baraja española de 40 cartas. Si uno de los casos aparece, queda excluido el otro.

$$P_1 = \frac{4}{40} = \frac{1}{10} \text{ (As)}$$

$$P_2 = \frac{4}{40} = \frac{1}{10} \text{ (Rey)}$$

$$P = P_1 + P_2 = \frac{1}{10} + \frac{1}{10} = \frac{2}{10} = \frac{1}{5}$$

2. Se dice que dos o más sucesos son independientes, si la probabilidad de presentación de alguno de ellos queda influenciada por la presentación del otro. En caso contrario, se dice que son dependientes.

   En otras palabras, si el resultado de un suceso no afecta al otro, se dice que son independientes, por lo tanto se efectuará la multiplicación de las probabilidades para cada suceso.

   $P = P_1 \text{ x } P_2 \text{ x } P_3 ..... \text{ x } P_n$

   Ejemplo:

   ¿Qué probabilidad tendremos de obtener dos reyes, sacando una carta de una baraja y la otra de una segunda baraja?

$$P = \frac{4}{40} \text{ x } \frac{4}{40} = \frac{16}{1600} = \frac{1}{100}$$

   Se dice que los sucesos son dependientes, o eventos compuestos, si la ocurrencia o no ocurrencia de un evento en cualquier prueba afecta la probabilidad de otros eventos en otras pruebas, es decir que la probabilidad del segundo suceso depende del primer suceso, el del tercero, de lo que haya sucedido en el primero **y** segundo **y** así sucesivamente.

Ejemplo:

La probabilidad de obtener tres ases, sacando sucesivamente tres cartas de una baraja española, sin volverlas a incluir (sin reposición) en el montón.

$P_1 = \frac{4}{40} \qquad P_2 = \frac{3}{39} \qquad P_3 = \frac{2}{38}$

$$P = \frac{4}{40} \bullet \frac{3}{39} \bullet \frac{2}{38} = \frac{24}{59280}$$

$$P = \frac{1}{2470}$$

4.4 Análisis combinatorio

El análisis combinatorio es un sistema que permite agrupar y ordenar, en diversas formas, los elementos de un conjunto.

Los tres principales tipos de análisis combinatorio son:

4.4.1 *Permutaciones*

Se denominan permutaciones de η elementos, los diferentes grupos que se pueden hacer, tomándolos todos cada vez.

Las permutaciones implican orden.

Cada conjunto ordenado de η elementos se denominará una permutación de los n elementos diferentes.

La formula es $P_n = n!$, donde P_n corresponde al número de permutaciones posibles.

Por ejemplo:

Determine el número de permutaciones posibles de las letras A, B, C, D.

$P_4 = 4! = 4 \times 3 \times 2 \times 1 = 24$

Representémoslas:

ABCD	BACD	CABD	DABC
ABDC	BADC	CADB	DACB
ACBD	BCDA	CBAD	DBAC
ADBC	BDAC	CDAB	DCBA
ADCB	BDCA	CDBA	DCAB
ACDB	BCAD	CBDA	DBCA

Otro ejemplo:

¿Cuántos números diferentes de 9 cifras se pueden formar con los dígitos del cero al 9, usándolos una sola vez?

P_{10},= 10! = 10 x 9 x 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1 = 3.628.800

Permutaciones con repetición

Las permutaciones con repetición r, son un caso particular de las variaciones y no existe una ley sencilla para su formación. Dado lo complicado del sistema, sólo, se presenta la fórmula que logra el número de esta clase de permutaciones.

Sean los elementos aa - bbb - cc - d, para permutar con repetición, tendremos 8 elementos repartidos así: dos del primero, tres del segundo, dos del tercero y uno del cuarto, entonces las permutaciones se presentarán así:

P_8(r: 2,3,2,1) y la fórmula respectiva será:

$$P_8(r:2,3,2,1) = \frac{8!}{2!3!2!} = \frac{8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times 3 \times 2 \times 1 \times 2 \times 1}$$

P_8(r:2,3,2,1)=1680

La formula será $P_n(r: r^1, r^2) = \frac{n!}{r_1!r_2!}$

Ejemplo:

¿De cuántas maneras distribuiríamos 3 monedas de $100 y 4 monedas de $500 en una misma línea?

$$P_7(r: 3,4) = \frac{7!}{3!4!} = \frac{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 4 \times 3 \times 2 \times 1}$$

P_7(r: 3,4) = 35

Otro ejemplo:

Cuántos grupos de 2 letras se pueden formar con las letras de la palabra amigas.

$$P_6(r: 2) = \frac{6!}{2!} = 360$$

4.4.2 *Variaciones*

Las variaciones corresponden a aquellas permutaciones donde los elementos no se toman en su totalidad.

Dado un conjunto de n elementos diferentes, se denominará permutación parcial o variaciones, de subconjunto de r elementos (r<n) pertenecientes al conjunto dado.

La fórmula es: $V_r^n = \frac{n!}{(n - r)!}$

En el ejemplo de las cuatro letras o elementos (n) vamos a permutar de a 2 (r)

$$V_2^4 = \frac{4!}{(4 - 2)!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1} = 12$$

AB	BA	CA	DA
AC	BC	CB	DB
AD	BD	CD	DC

Ejemplo:

¿Cuántas cifras diferentes de 4 dígitos se pueden formar con los dígitos del 0 al 9, usándolos una vez?

$$V_4^{10} = \frac{10!}{(10 - 4)!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 5.040$$

4.4.3 *Combinaciones*

Son aquellas en las que no interesa el orden de la aparición de elementos del conjunto. Será lo mismo AB que BA.

Cuando se toma la totalidad de elementos, solamente se puede hacer una combinación. Si se tiene las letras A, B, C, D y se combinan de dos en dos, se obtienen 6 casos:

AB	BC	CD
AC	BD	
AD		

La fórmula será:

$$\binom{n}{r} = C_r^n = \frac{n!}{r!(n - r)!}$$

Y se lee combinación de n elementos tomados de r en r.

Ejemplo:

Cuál es la combinación de las 4 letras A, B, C, D, tomadas de 4 en 4.

$$\binom{4}{4} = C_4^4 = \frac{4!}{4!(4 - 4)!} = \frac{4 \times 3 \times 2 \times 1}{4 \times 3 \times 2 \times 1 \times 1} = 1$$

El factorial de cero es uno (0! = l).

En la combinación de estas 4 letras tomadas de 2 en 2 será,

$$\binom{4}{2} = C_2^4 = \frac{4!}{2!(4 - 2)!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1} = 6$$

4.5 Probabilidad condicional

En algunos eventos probabilísticos se tiene alguna información, que reduce el espacio muestral original a uno de sus subconjuntos; las probabilidades asociadas a esos subconjuntos determinan la probabilidad condicional.

Su fórmula está dada así:

$$P(B/A) = \frac{P(B \cap A)}{P(A)}$$

Ejemplo:

El 18 % de las familias de un barrio tienen vehículo propio, el 20% tiene vivienda de su propiedad y el 12% tiene vivienda y vehículo. ¿Cuál es la posibilidad de tener vivienda si se tiene vehículo?

A = Propietario de vehículo

A’ = No propietario de vehículo

B Propietario de vivienda

B’ No propietario de vivienda

	B	B’	Total
A	0.12	0.06	0.18
A’	0.08	0.74	0.82
Total	0.20	0.80	1.00

$$P(B/A) = \frac{0.12}{0.18} = 0.66$$

INTRODUCCIÓN A LAS PROBABILIDADES
AUTOEVALUACIÓN N° 3

1. La definición clásica de probabilidades es el cociente de dividir el número de casos favorables que pueden ocurrir en un suceso, por el número de casos posibles: Sí ______ No ______

2. ¿Cuál es la probabilidad de que una bola, extraída al azar de una urna que contiene 3 bolas rojas, 4 blancas y 5 azules, sea blanca?

   a. 3/12

   b. 1/3

   c. 5/16

   d. Ninguna

3. Una señora invita a cenar a 8 personas; después de sentarse ella, ¿de cuántas maneras se pueden sentar sus invitados?

4. En un juego de moneda, entre dos personas, con un premio de $1.000 por aparición de cara, ¿cuál es la esperanza de ganar con el resultado de cara?

5. ¿De cuántas maneras se pueden sacar dos manzanas de una caja que contiene 8 manzanas?

RESPUESTAS A LA AUTOEVALUACIÓN N° 3

1. Sí.
2. 1/3
3. 40.320
4. $500
5. $C^8_2 = 28$

4.6 Distribución binomial

En repetidos ensayos, siendo **p** la probabilidad de éxito en un solo ensayo la cual debe permanecer fija y **q** la probabilidad de fracaso, entonces la probabilidad **P** de que se obtengan **r** éxitos en **n** ensayos, es el termino del desarrollo binomial $(q+p)^n$.

La fórmula general será:

$$p = C^n_r \, p^r \, q^{n-r}$$

Los criterios que deben satisfacer una experiencia binomial son:

a) Existir un número fijo de pruebas repetidas (n).

b) Cada una de las **n** pruebas debe tener dos resultados: favorable o desfavorable; en el caso de una moneda será: cara o sello.

c) La probabilidad de éxito de un acontecimiento es fija.

d) Las pruebas son independientes.

e) Nos interesa el número de éxitos en **n** pruebas.

En caso de lanzar 4 monedas y se quiera determinar la probabilidad de obtener exactamente dos caras, será:

$$P = C^4_2 \left\{\frac{1}{2}\right\}^2 \left\{\frac{1}{2}\right\}^{4-2} = \frac{4!}{2!2!} \left\{\frac{1}{2}\right\}^2 \left\{\frac{1}{2}\right\}^2 = \frac{4.3.2.1}{2.1.2.1} \left\{\frac{1}{4}\right\} \left\{\frac{1}{4}\right\}$$

$$P = 6 \left\{\frac{1}{16}\right\} = \left\{\frac{6}{16}\right\} = 0.375 = 37.5\%$$

4.7 Distribución normal

Muchas distribuciones de mediciones, que se hacen tanto en las ciencias sociales como en las ciencias naturales, tienden a tener un polígono de frecuencias con una forma que se asemeja al corte transversal de una campana.

Esta distribución se observa más cuando el número de observaciones es grande y cuando en muchos casos las investigaciones se realizan con muestras de poblaciones grandes; en la mayoría de los casos las distri-

buciones tienden a aproximarse a la curva en forma de campana ya mencionada.

Una distribución de frecuencias que se asemeja a una campana se reconoce como una Distribución Normal, que por tener características especiales, se convierte en un requisito fundamental para entender el proceso relacionado con la inferencia estadística y la prueba de hipótesis.

Cuando el número **n** de experimentos es grande, se hace difícil el cálculo de las probabilidades mediante el teorema del Binomio; para facilitar el cálculo y agilizar las operaciones, utilizaremos la distribución normal.

Entre las propiedades de la distribución normal están,

a. El área total debajo de la curva, puede representar el número total de elementos en una población y se puede considerar que el área es igual a la unidad.

   Esta propiedad permite determinar la proporción de la frecuencia contenida entre dos puntajes a lo largo del eje horizontal, ya que se puede calcular la proporción del área debajo de la curva y contenida entre los dos puntos.

b. La media aritmética, la mediana, y la moda de una distribución normal son iguales a cero.

c. La desviación estándar es igual a 1

d. Como veíamos, tiene una distribución simétrica.

e. Es una distribución asintótica, es decir, como no toca el eje de las **X,** se extiende infinitamente.

f. En una distribución normal el 68.26% de los valores bajo la curva, se encuentra dentro de más o menos una desviación estándar de la media aritmética; el 95.46% de los valores debajo de la curva se encuentra dentro de más o menos dos desviaciones estándar de la media y el 99.9% se encuentra dentro de más o menos tres desviaciones estándar, tal como se ve en la gráfica 5.

La curva está dada por la función:

$$Y = \frac{n}{G\sqrt{2\pi}} \, e^{\frac{z^2}{2G^2}}$$

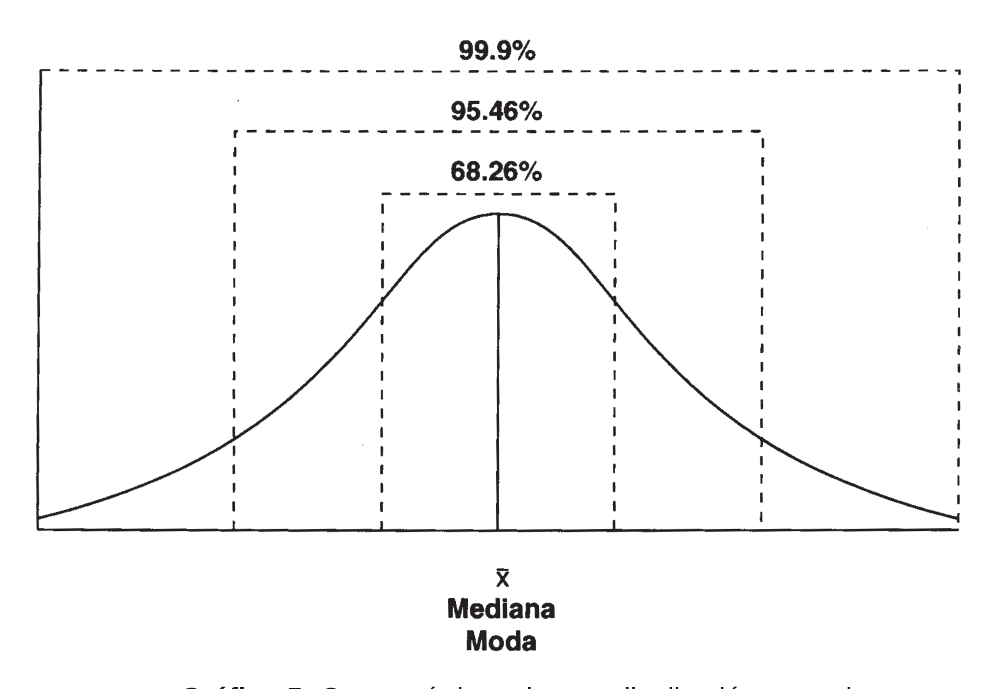
donde: n número de datos

G desviación estándar de la distribución binomial $= \sqrt{npq}$

e base de los logaritmos naturales = 2.71828

π 3.1416

µ media de la distribución nominal = np



Gráfica 5. Características de una distribución normal.

Probabilidad mediante una aproximación a la curva normal

X y **Y** son cantidades concretas de la misma especie. La variante estadística $z = \frac{\chi - \mu}{G}$, representa una medida de las desviaciones estándar o de las G llamadas unidades estandarizadas. La expresión anterior también es conocida como desviación normal. Aplicando la fórmula podemos calcular la probabilidad, por ejemplo de obtener 4, 5 y 6 caras, en 9 lanzamientos de una moneda, mediante una aproximación a la binomial.

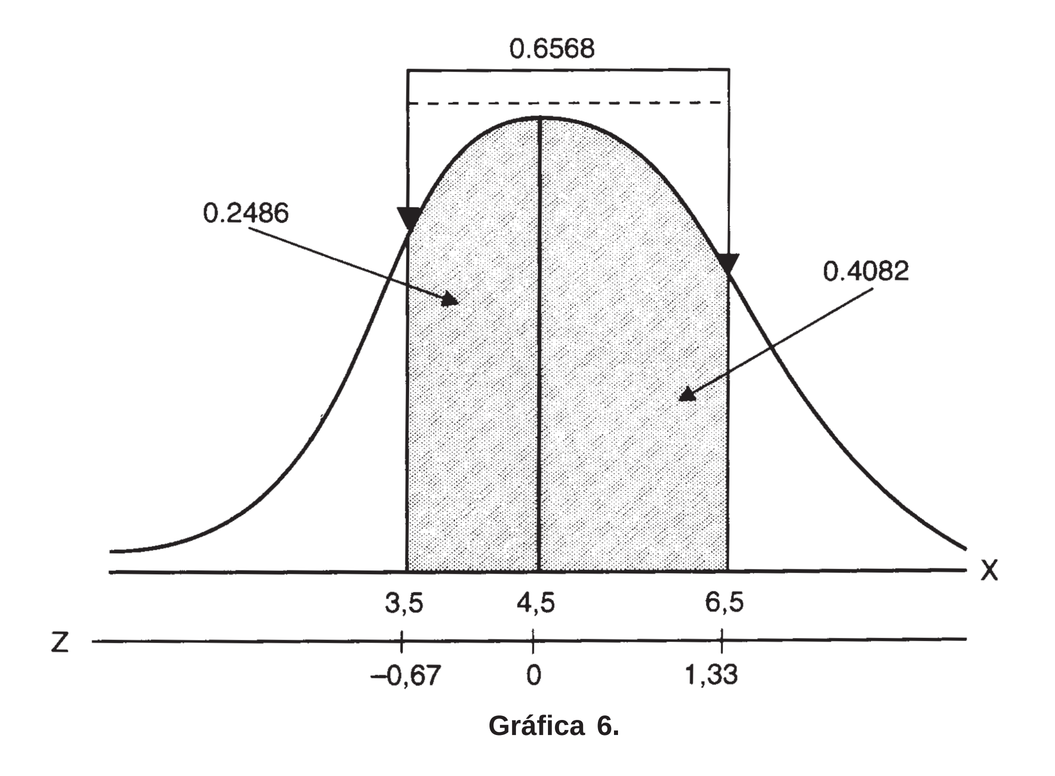
$n = 9 \quad p = ½ \quad q = ½$

$\mu = np = 9\,(½) = 4.5$

$G = \sqrt{npq} = \sqrt{9\,(½)\,(½)} = 1.5$

$p = (3.5 < x < 6.5)$

x = 4,5 y 6 caras



Gráfica 6.

Se tiene que x 3.5 corresponde al límite inferior de 4, y x 6.5 es el límite superior de 6. Se va a buscar el valor del área comprendido entre 3.5 y 6.5, o sea, P (3.5<x<6.5). Primero localizamos el área a partir de la media hasta el límite inferior, dado que cada lado vale el 50%, la suma total será igual a uno.

De donde $z = \dfrac{x - \mu}{G}$ entonces $z = \dfrac{3.5 - 4.5}{1.5} = -0.67$

z = (-0.67) = 0.2486, tomado de la tabla, área bajo la curva normal.

$$Z = \frac{6.5 - 4.5}{1.5} = 1.33$$

Z = 1.33 = 0.4082, tomado de la tabla, área bajo la curva normal.

Se desea el área comprendida entre z = -0.67 y z = 1.33

Para ello sumamos: 0.2486 + 0.4082 = 0.6568

La probabilidad de que aparezca 4, 5 y 6 caras es del 65.68%. Véase gráfica 6.

Ejemplos:

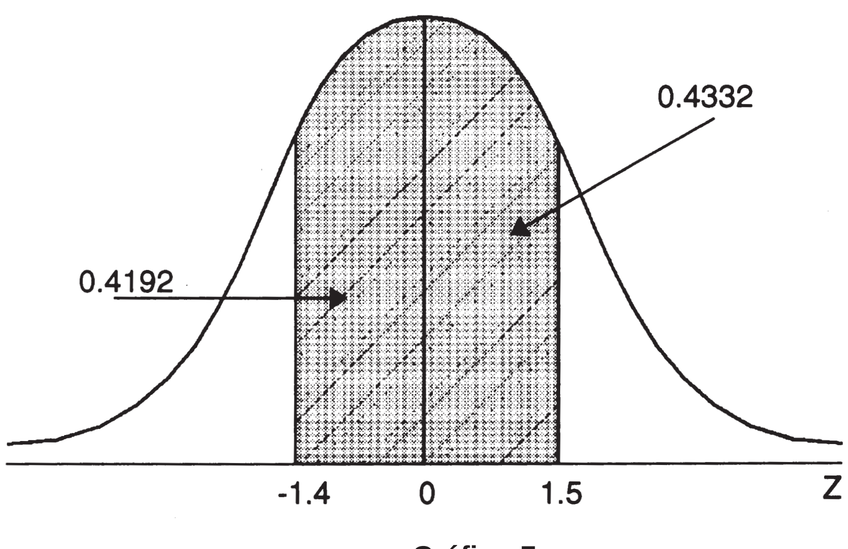
Usando la tabla del área bajo la curva normal, plantearemos algunos problemas que servirán de modelo. Ver gráficas: 7 - 8 - 9 y 1 0.

a) Para z = 1.5 y z = -1.4

z = 1.5 la tabla da 0.4332 z =1.4 la tabla de 0.4192

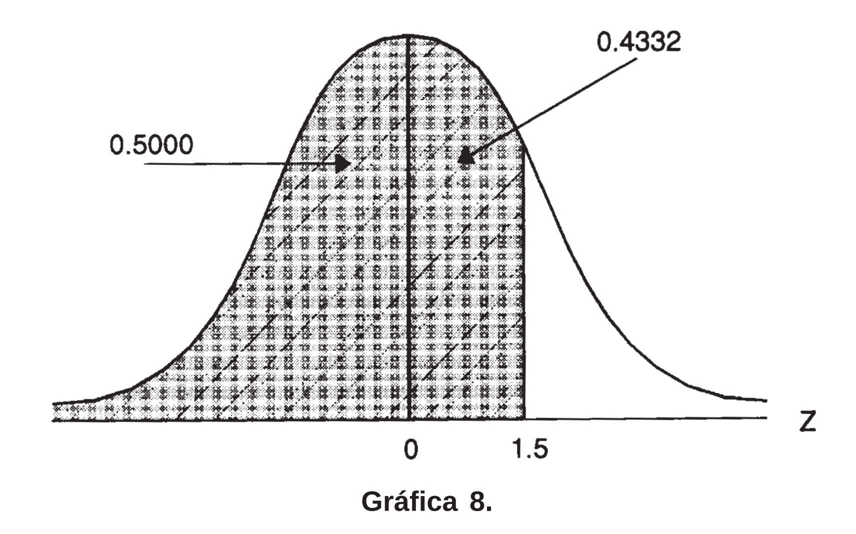
p (-14<z<1.5) = 0.4332+0.4192

p 0.8524 = 85.24%



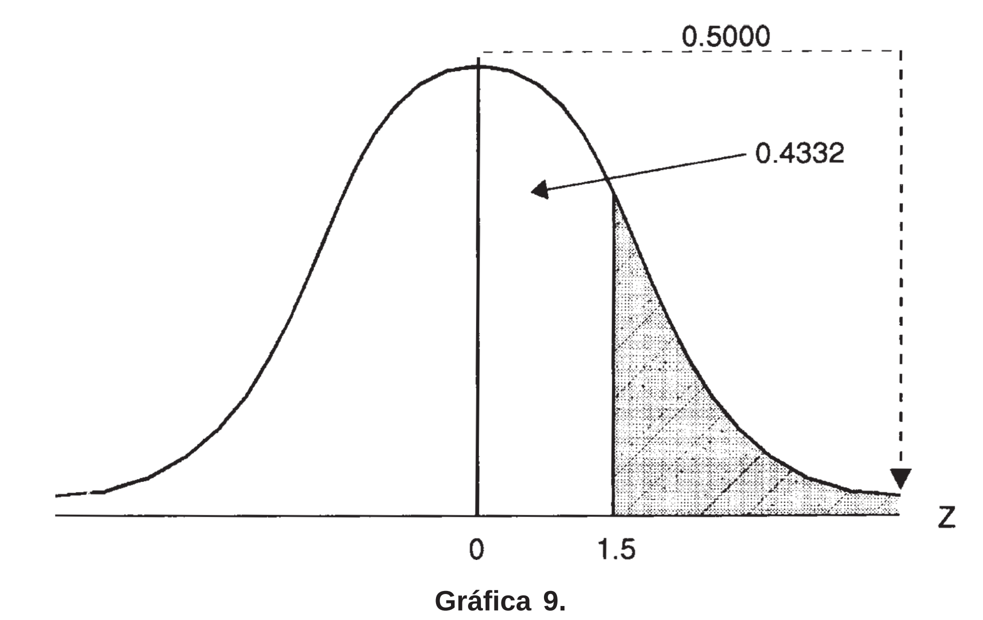
Gráfica 7.

b) Si se plantea p (z<1.5), corresponde al siguiente gráfico.
0.5000 + 0.4332 = 0.9332 = 93.32%



Gráfica 8.

c) P (z>1.5)
0.5000 - 0.4332 = 0.0668 =6.68%



Gráfica 9.

d) Si queremos conocer el porcentaje de frecuencia comprendida entre

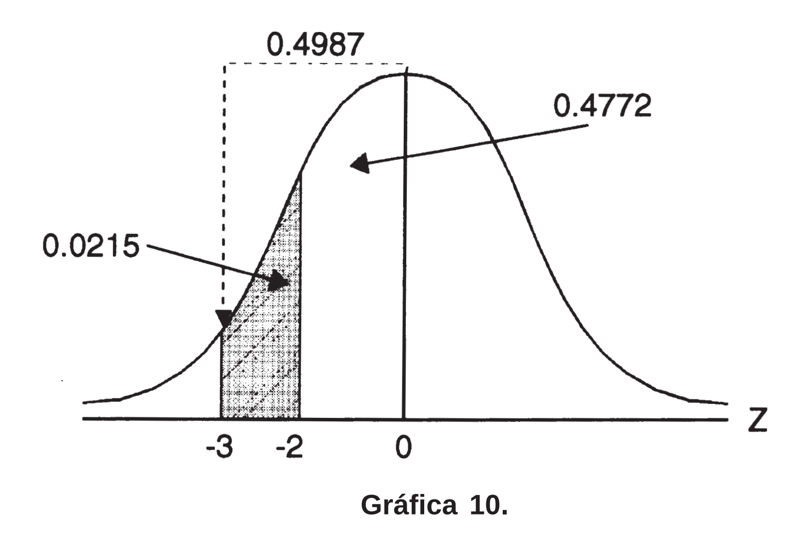
$z = -3$ y $z = -2$, será necesario proceder por diferencia.

$z = -3 = 0.4987$

$$z = -2 = \frac{0.4772}{0.0215}$$

La diferencia es 0.0215 = 2.15%

$P(-3 \leq z \leq -2) = 2.15\%$



Gráfica 10.

4.8 Distribución de Poisson

Se tiene ahora una distribución binomial, cuando n es grande, por lo general mayor de 50, a la vez que **p,** la probabilidad de éxito de un suceso se acerque a cero, mientras que q, la probabilidad de fracaso, se aproxime a uno, de tal manera que el producto **np,** llamado lamda λ es menor o igual a 5, debe utilizarse la distribución de Poisson.

Su fórmula es:

$$P = \frac{\lambda^x e^{-\lambda}}{X!}$$

donde e = 2.71828

$\lambda = np$

x = número de casos favorables

Ejemplo:

a) Si el 1 % de las bombillas fabricadas por una compañía son defectuosas, hallar la probabilidad de que en una muestra de 100 bombillas 3 sean defectuosas.

$\lambda = 100 \times 0.01 = 1$
$X = 3$

$$P = \frac{1^3 - e^{-1}}{3!} = \frac{1\ (0.36788)}{3.2.1} = 0.06131$$

e^{-1} (utilizamos la tabla para valores de $e^{-\lambda}$)

$e^{-1} = 0.36788$

b) Si, la probabilidad de que una persona adquiera la enfermedad como consecuencia de una vacuna contra la misma, es 0.0002, ¿cuál es la probabilidad de que la adquieran exactamente 5 personas en una población de 10.000 vacunados?

$$p = \frac{\lambda^x \cdot e^{-\lambda}}{X!}$$

$\lambda = np = 10.000 \times 0.0002 = 2$
$x = 5$

$$\text{entonces } P = \frac{2^5\ (0.13534)}{5.4.3.2.1} = \frac{32(0.13534)}{120} = \frac{4.33088}{120}$$

P = 0.03609 = 3.61 %

P (x = 5) = 3.61 %

LA CURVA NORMAL Y SUS APLICACIONES
AUTOEVALUACIÓN N° 4

1. La curva normal se caracteriza por que es asimptótica. Esto quiere decir que:

   a. La media es igual a la moda __________________

   b. Tiene forma de campana __________________

   c. Es simétrica __________________

   d. Nunca toca el eje __________________

   e. Ninguna de las anteriores __________________

2. La tabla bajo la curva normal sirve para indicar cuál es:

   a. La proporción de los casos que se encuentran entre la media y un valor llamado puntaje ______________

   b. El valor de la desviación estándar en relación con el valor de la media ______________

3. En una distribución normal, la media, la mediana y la moda son iguales:

   Sí _____ No _____

4. ¿Cuál es el área bajo la curva normal que corresponde a un z = 2.5?
   ____

5. El promedio de las notas de un curso está en 1.350 puntos y tiene una distribución normal con una desviación estándar de 18 puntos.

   a. ¿Qué porcentaje de alumnos tiene notas entre 1.350 y 1.377 puntos? __________________

   b. Entre 1.365 y 1.377 __________________

   c. Entre 1.31 y 1.351. __________________

6. Con una distribución de 20 elementos se obtiene una media de 20.3 y una desviación estándar de 2.81. Si uno de los elementos tiene un valor de 28, ¿cuál es el valor de z?

   a. 17.18 ______________

   b. 7.70 ______________

   c. 2.74 ______________

   d. 5.48 ______________

   e. Ninguno de los anteriores ______________

7. En una distribución normal, el 68.26% de los valores bajo la curva se encuentra dentro de más o menos una desviación estándar de la media aritmética.

   Sí _____ No _____

8. En una distribución binomial, cuando n es pequeña, se utiliza la distribución de Poisson.

   Sí _____ No _____

RESPUESTAS A LA AUTOEVALUACIÓN N° 3

1 =d; 2=a-, 3=Sí; 4=49.38%; 5=a)43.32%;

b)13.65; c)21.14%; 6=c)2.74; 7=Sí; 8=No

5. ESTADÍSTICA INFERENCIAL

Ayuda a determinar la confiabilidad de la inferencia de que los fenómenos observados en la muestra ocurrirán también en la población de donde se seleccionó la muestra. Es decir, sirve para estimar la eficacia del razonamiento inductivo con el cual se infiere que lo que se observa en una parte se observará en el grupo entero.

El objeto de la estadística es el de hacer inferencias (predecir, decidir) sobre algunas características de la población con base en la información contenida en una muestra.

Por lo tanto, se puede afirmar que el empleo de la estadística inferencial presume el dominio de la estadística descriptiva. Sin embargo se debe tener presente que al predecir un resultado o tomar una decisión se está sujeto a una incertidumbre o margen de error, bien sea por la técnica de muestreo o por la selección del estadístico de prueba.

Es importante resaltar que se debe diferenciar entre las técnicas que son válidas para el análisis de los datos cualitativos y de los datos cuantitativos.

En el caso de la estadística inferencial el investigador se encuentra con dos tipos de técnicas. Por un lado las técnicas **paramétricas** que suponen una serie de supuestos acerca de la naturaleza de la población de la que se extrajo la muestra de estudio. Por otro lado, las técnicas **no paramétricas,** que no requieren supuestos sobre las características de la población y que se facilitan más para el análisis de datos nominales y ordinales.

Las ventajas de las técnicas paramétricas es que son más potentes que las no paramétricas y por consiguiente las inferencias que se realizan son más fiables. El inconveniente es que el investigador no siempre puede cumplir con los requisitos y supuestos que exige el enfoque paramétrico, sobre todo en investigaciones educativas y sociales.

La ventaja de las técnicas no paramétricas es que son fáciles de utilizar y algunas son tan potentes como las paramétricas.

El análisis de datos cualitativos ha generado técnicas propias, que actualmente constituyen toda una metodología específica que viene marcada por la propia idiosincrasia cualitativa y que toma determinadas opciones en relación a las unidades de registro de los datos y la forma de tratarlos.

Actualmente existe interés por sistematizar estos procedimientos, sin embargo esta tarea es difícil debido al carácter abierto y flexible de la metodología, los diversos enfoques, y la variedad de objetivos científicos que puede cubrir.

Una de las maneras de realizar estimaciones es con las pruebas de hipótesis referentes a sus valores.

En muchos aspectos el procedimiento formal para la prueba de hipótesis es similar al método científico.

5.1 La prueba de hipótesis

Ya el investigador, desde el momento en que identifica y delimita el problema que va a investigar, ha aclarado los elementos conceptuales y las hipótesis o modelos que tiene que probar. En las etapas complicadas con la generación de los instrumentos, con la recolección de la información y con la codificación de la misma, estas hipótesis se aclaran mucho más y en algunos casos, toman características cuantitativas, permitiendo la aplicación de técnicas estadísticas, para probar su relevancia en cuanto a su capacidad para generalizar el fenómeno, o la relación entre fenómenos, a poblaciones mayores, o para tomar decisiones en relación con tales fenómenos.

En esta parte distinguiremos la lógica del proceso para probar una hipótesis y aplicaremos en casos determinados la prueba Z, la prueba T y la prueba de significación entre muestras, para ilustrar el proceso de prueba de hipótesis.

La lógica de la prueba de hipótesis

En el proceso de la estadística inferencial, hay dos tipos de hipótesis claves: una es la hipótesis de investigación (H_1), que simplemente señala la existencia de un hecho o de un evento, o la relación entre dos o más fenómenos.

La otra es la hipótesis nula (H_0) que se construye artificialmente para que el investigador evalúe su hipótesis de investigación.

Si la hipótesis de investigación afirma que existe una relación entre dos o más fenómenos, la hipótesis nula (H_0) plantea que no existe relación entre los dos fenómenos.

Es decir, se utiliza esta última como un modelo para probar la hipótesis de investigación, o sea, probar su veracidad o su falsedad.

Para probar una hipótesis, se necesita:

1° La existencia o planteamiento de la misma.

2° Que se anticipen los eventos que pueden resultar a partir de la observación o experimentación.

3° Que se especifiquen las condiciones bajo las cuales la hipótesis se puede rechazar o aceptar.

4° Que se haga el experimento y la observación y que se analice el evento o resultado.

5° Que se tome la decisión de rechazar o aceptar la hipótesis.

Esto implica que, aunque se tiene que observar o experimentar, todas las suposiciones se hacen antes de las mismas observaciones o experimentos. Es decir, el investigador predice y luego bajo las reglas del juego de su predicción evalúa en relación con los resultados. Si éstos no son consistentes con lo que se predijo, se rechaza la hipótesis.

Con un ejemplo podemos ilustrar la naturaleza de la prueba de hipótesis. Supongamos que se está probando una nueva semilla para resolver problemas de productividad en una región determinada. La hipótesis de la investigación sería que la media aritmética por fanegada es mayor de 519 arrobas, que es el tradicionalmente observado en esa región.

Para verificar la hipótesis tenemos que completar los cinco pasos siguientes:

1° Aclarar la hipótesis de investigación H_1 y crear la hipótesis nula H_0.

2° Obtener la distribución del muestreo.

3° Seleccionar un nivel de significancia y una región de rechazo.

4° Hacer la prueba estadística.

5° Comparar y concluir.

Punto 1

El paso uno se define en el enunciado del problema. Para que sea verificada H_I, el proceso a seguir es contradecir a H_0, es decir, hay que contradecir que la media aritmética es igual o menor a 519.

Entonces las dos hipótesis se representan así:

H_1: X $\geq$ 519
H_0: X $\leq$ 519

Con lo anterior, lo que se está diciendo es que, si encontramos que el promedio de producción por fanegada es de 519 arrobas o menos, tendremos que concluir que nuestra hipótesis inicial H_1, tendrá que ser modificada y por lo tanto, la nueva semilla, por lo menos en cuanto a su productividad, no es competitiva con las tradicionalmente utilizadas.

Punto 2

Una vez aclaradas las suposiciones del punto 1, acerca de H_1 y H_0, por medio de un razonamiento matemático podemos obtener la distribución de muestreo, o sea, la distribución de todos los posibles valores que una estadística puede tener, con base en un muestreo probabilístico de un universo.

Esta distribución nos dice, entonces, qué probabilidad hay de obtener cada uno de los resultados.

Dependiendo del tamaño de las muestras, las distribuciones de muestreo tienden a tomar una distribución normal y por ello no es necesario calcularlas.

Un teorema importante, que es una aplicación de la distribución normal, conocido como el teorema del Límite Central, dice que si en una población de tamaño N, con una media de μ y una varianza de S^2 se obtienen muestras al azar, la distribución de las medias de las muestras seleccionadas será normal -y más lo será en la medida en que se incremento el número de muestras seleccionadas- y tendrá una media de Y y varianza S^2/N.

El teorema garantiza que la media de una muestra puede estar muy cercana al de la población, en la medida en que tenga un tamaño **n** grande, ya que la varianza de la distribución de las medias muestrales S^2/N, se hace más pequeña en la medida en que aumenta el tamaño **n**.

La desviación estándar de la distribución de las medias de la muestra se denomina como el error estándar: $S/ /N$

Punto 3

A partir de la información obtenida, el investigador puede escoger resultados que permitan aceptar o rechazar sus supuestos. Posteriormente, los resultados identificados para rechazar la suposición se llaman de región crítica, dividiendo así los resultados en los que sirven para rechazar y los que sirven para aceptar la hipótesis nula, sabiendo que estos sectores están relacionados con errores de tipo α y β.

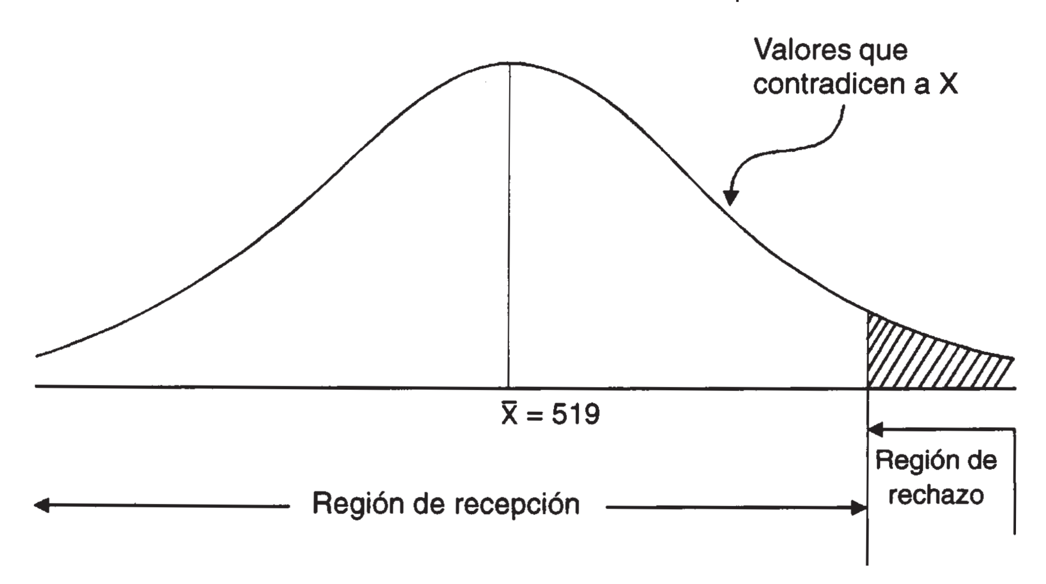
Una vez aclaradas la H_1 y H_0, obtenemos aleatoriamente una muestra de la producción por fanegada de la semilla probada en la región.

Calculamos la media aritmética y la desviación estándar de los datos de la muestra, información ésta que nos permite hacer la prueba estadística.

Si calculamos $\bar{x}$, sabemos que la distribución de muestreo de x, suponiendo que H_0 es cierto, es bastante aproximada a una distribución normal con un $\mu = 519$.

Valores de $\bar{x}$, que contradicen H_0 y por lo tanto validan a H_1, serán aquellos que se encuentran en la región de la cola derecha de la distribución, como lo señala la figura 11. Estos valores contradictorios generan la región de rechazo, es decir, si el valor observado de x cae en la región de rechazo, descartamos H_0 y aprobamos H_1. Observemos que se verifica H_1 contradiciendo H_0. Por lo tanto si x cae en la región de aceptación de la figura 11, aceptamos H_0 y no podemos verificar H_1. Esta es la etapa de comparación con la región de rechazo.

Gráfica 11. Suponiendo que H_0 es verdadera, los valores contradictorios de X están en la cola superior.



Hay posibilidad de cometer dos tipos de errores al probar la hipótesis. Podemos, como en cualquier proceso de decisión entre dos escogencias, cometer el error de rechazar a H_0 cuando en realidad es verdadera; error tipo I ó α; o podemos aceptar H_0 cuando en realidad es falsa; error tipo II ó β. Y aunque es conveniente determinar la región de aceptación y de rechazo, minimizando tanto el error tipo α como el error tipo β, esto no es imposible porque están inversamente relacionadas, es decir, para una muestra de tamaño η, cuando se cambia la región de rechazo para incrementar el error α, el error β disminuye y viceversa.

Para resolver este problema, la comunidad de científicos ha decidido que se debe trabajar con la probabilidad de cometer error tipo I ó α y la especificación del valor para el error tipo α determina la región de rechazo.

¿Cómo se hace? Volviendo a nuestro ejemplo, sabemos que rechazamos H_0 si obtenemos grandes valores de x.

Si tenemos una muestra de 36 terrenos de una fanegada, como muestra del estudio y si la media aritmética es $\bar{x} = 573$ y la desviación estándar es S = 124 la pregunta clave sobre la cual tenemos que tomar una decisión es si el $\bar{x}$ para toda la región es mayor que 519.

Para concluir, decidimos el valor del error αcon el cual queremos trabajar; siempre sabemos que hay la posibilidad de cometer un error. Si en este caso queremos correr el riesgo de equivocarnos 1 de cada 40 decisiones, es decir de rechazar incorrectamente H_0 entonces $\alpha = 1/40 = 0.025$

Como suponemos que H_0 es verdadera, x está distribuida normalmente y tiene μ = 51 9 y una desviación estándar de:

$$\frac{S}{\sqrt{N}} = \frac{124}{\sqrt{36}} = 20.67$$

Punto 4

Supongamos que en la figura 11 la región rayada sea la equivalente a la de un α = 0,025. Para calcular el punto de rechazo, se tiene que calcular el valor de Z, que tiene un área de 0.025 a su derecha. En la tabla de áreas bajo la curva normal vemos que el valor correspondiente es de 1.96, por lo que lo designaremos como el límite para la región de rechazo y si el valor observado de x es mayor que 1.96, desviaciones estándar arriba de μ= 159, se rechaza la hipótesis nula.

Cuando el valor de $\bar{x}$ es igual a 573 y el de S es igual a 124, el valor de Z normalizado es:

$$Z = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N}}} = \frac{573 - 519}{\frac{124}{\sqrt{36}}} = \frac{54}{20.67} = 2.61$$

Punto 5

El último paso es comparar el valor obtenido con el dispuesto como rechazo, ya que $\bar{x}$ encontrado está a más de 1.96 unidades de desviación estándar arriba del $\mu = 519$, rechazamos H_0 en favor de H_1 y concluimos que la productividad de la nueva variedad de semilla es significativamente mayor que 519.

5.2 Pruebas de significancia de muestras únicas o simples

Cuando se quiere probar hipótesis a partir de una muestra única, se pueden utilizar varias pruebas de significancia de hipótesis. En estudios en ciencias naturales y ciencias sociales es muy frecuente este tipo de situaciones y las técnicas utilizadas más frecuentemente son la prueba Z, y la prueba t frecuentemente llamada t de Student.

5.2.1 *Prueba Z*

Normalmente se utiliza cuando se conoce el parámetro de la población. Ilustremos su uso con el siguiente ejemplo hipotético:

En una institución de enseñanza media se presenta una deserción que los directores consideran muy alta y pensaron que con un programa de consejería, podrían controlarla, con la hipótesis de que cuanta más guía personal, mayor sería la probabilidad de continuar exitosamente en la institución. Se generó una consejería y al cabo de un año se recopiló información sobre el número de veces que cada estudiante tuvo contacto formal con un consejero.

La media aritmética para el total de estudiantes fue de 12.1 visitas al año, con una desviación estándar de 3.21.

Meses después se hizo una muestra aleatoria simple de 40 estudiantes desertores y a partir de los datos obtenidos con la recopilación de información, se les calculó una media aritmética de 9.11. Al ver que era más baja que el resto de la población, los directores se hicieron la siguiente pre-

gunta. ¿Es el número de visitas de los desertores significativamente más bajo que el de la población estudiantil?

Supongamos la hipótesis de la investigación H_1: $\bar{x} < \mu$ y la hipótesis nula H_0: $x \geq \mu$, decidimos utilizar un nivel del 0.01 de significancia, la muestra, como vimos, es probabilística.

Para probar la hipótesis calculamos el puntaje Z, normalizado como habíamos visto así:

$$Z = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N}}} = \frac{9.11 - 12.1}{\frac{3.21}{\sqrt{40}}} = \frac{-2.99}{\frac{3.21}{6.32}} = -5.86$$

El signo es indiferente al resultado, porque estamos trabajando con áreas, por lo tanto Z = 5.86.

En la tabla de áreas bajo la curva normal, vemos que para que el $\bar{x}$ sea significativamente distinto de $\mu = 12.1$, se necesita una Z = 2.33 o mayor; como el Z obtenido es mayor, se rechaza H_0 y se acepta H_1, ya que:

0.5000 - 0.0100 = 0.4900

Entonces Z = 2.33

Concluimos que los estudiantes desertores tienen un número de contactos significativamente inferior al resto de la población estudiantil.

5.2.2 *Distribución t de Student*

En los casos en que no se conoce el parámetro poblacional, dependemos únicamente de las estadísticas de las muestras y utilizamos la Prueba t.

Cuando n > 30 (muestra grande) la desviación estándar simbolizada por $\mathcal{s}$, se le considera como un buen estimador de la desviación estándar poblacional, debido a que existe una mayor probabilidad de que los valores extremos que toma la variable, queden incluidos en el cálculo de la varianza para la muestra, tal como ocurre en el cálculo de la varianza poblacional.

Siendo, $\mathcal{s}$ = S, la fórmula para obtener $\mathcal{s}$ será:

$$\hat{s} = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$$

Para n $\leq$ 30 (muestra pequeña) la desviación estándar se simboliza por s, cuando no se ha efectuado ninguna corrección. Se considera que, s, por lo general, es mejor que S, debido a la menor probabilidad de que se incluyan los valores extremos de la variable. Por tanto, se hace necesario efectuar algunas correcciones en el cálculo:

a. Cuando se da la desviación estándar sin corregir:

$$\hat{s} = \sqrt{\frac{\sum(x - \bar{x})^2}{n}} \sqrt{\frac{n}{n - 1}}$$

b. Cuando se desea corregirla directamente

$$\hat{s} = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

En la distribución de t Student, se considera que las curvas son simétricas, pero algo más achatadas y más abiertas en los extremos, los cuales corresponden a regiones críticas. A medida en que el tamaño de la muestra se hace más grande, más se acerca a la normal.

La función dada para este tipo de distribución es:

$$Y = C\left\{1 + \frac{t^2}{v}\right\}^{-\frac{v+1}{2}}$$

donde v corresponde a los grados de libertad, número que depende de n; C es una constante que depende de v y es calculada en tal forma que el área bajo la curva sea igual a 1.

Grados de libertad

Corresponden al número máximo de variables que puedan asignarse libremente, antes de que el resto de las variables queden completamente determinadas.

Las variantes estadísticas para el cálculo de t son:

a) En las distribuciones muestrales donde $n \leq 30$, la fórmula es:

$$t = \frac{\bar{x} - \mu}{\frac{\hat{s}}{\sqrt{n-1}}}$$

también puede escribirse en la siguiente forma

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

donde s se obtiene corrigiéndola así:

$$s = \hat{s}\sqrt{\frac{n}{n-1}} \quad ó \quad s = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}}$$

b) En las distribuciones de medias muestrales se tiene:

$$t = \frac{(\bar{X} - \bar{Y})^2 - (\mu_x - \mu_y)}{s_{\bar{x}-\bar{y}}}$$

siendo

$$s^2 = \sqrt{\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_1 + n_2 - 2}}$$

$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$$

En la determinación de los puntos críticos, t_i (inferior) y t_s (superior), se utiliza la tabla t de Student. En primer lugar se fija el nivel de significación, por ejemplo, $\alpha = 0.05$, luego se calculan los grados de libertad, siendo en distribuciones muestrales $v = n - 1$, en diferencias de medidas muestrales: $v = n_i + n_2 - 2$. Si la prueba es bilateral, se tomará el 5% para cada una de

las regiones críticas y si es unilateral se tomará el doble del nivel de significancia asignado; en este caso, será 0. 1 0.

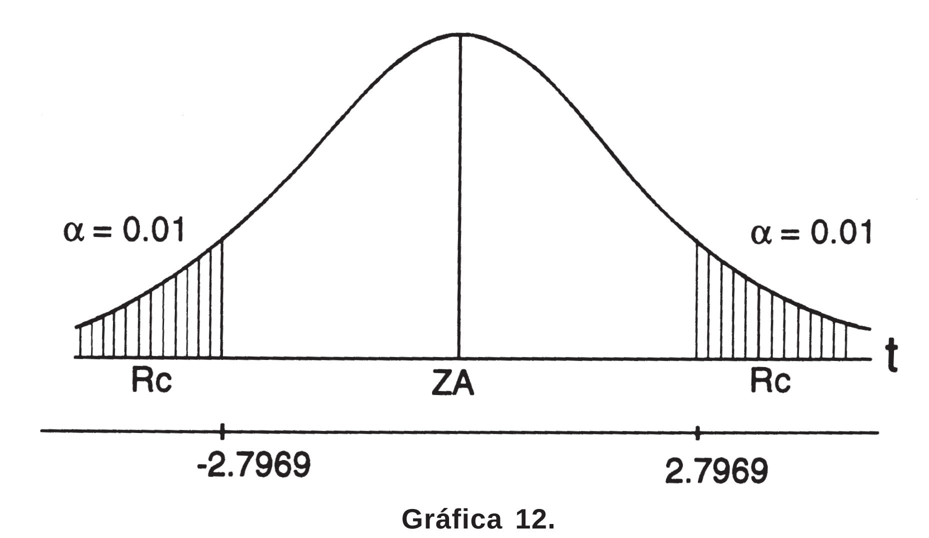
Ejemplos:

1. Una muestra de 25 observaciones tiene una media de 42.0 y una desviación estándar de 8. Trabajando con un nivel de significancia del 1%. Existe razón para rechazar la hipótesis de que la media de la población es de 46.0.

a. $H_1: \mu_x \neq 46$

$H_0: \mu = 46$

b. $\alpha = 0.01$



Gráfica 12.

c. $\hat{s} = 8$

$$\hat{s} = 8\sqrt{\frac{25}{25-1}} = 8(1.021) = 8.17$$

$$t = \frac{42 - 46}{\frac{8.17}{\sqrt{25}}} \qquad t = \frac{-4}{1.63} = -2.45$$

$V = 25\text{-}1$; $\alpha = 0.01$; corresponde a una $t = 2.7969$

Como el valor de t = -2.45 queda en la región de aceptación, se acepta la hipótesis de que $\mu = 46$, es decir no existe razón para rechazar que la media de la población es 46.

2. Un fabricante de cigarrillos analiza el tabaco de dos marcas diferentes, para determinar el contenido de nicotina y obtiene los siguientes resultados, en miligramos.

Marca A: 24 26 25 22 23
Marca B: 27 28 25 29 26

¿Los resultados anteriores, señalan que existe una diferencia en el contenido medio de nicotina en ambas marcas?

Solución:

a) $H_I: \mu_x \neq \mu_y$
$H_O: \mu_x = \mu_y$

b) = 0.05

c) $s_{\bar{x}-\bar{y}} = 1$

Siendo:

$$s = \sqrt{\frac{\sum(X - \bar{X})^2 + \sum(Y - \bar{Y})^2}{n_1+n_2-2}}$$

$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$$

Tabla 11. Resultado de las dos marcas de cigarrillos para determinar el contenido de nicotina

X	y	**x-$\bar{x}$**	**(x - $\bar{x}$)²**	y-$\bar{y}$	(y - $\bar{y}$)²
24	27	0	0	0	0
26	28	2	4	1	1
25	25	1	1	-2	4
22	29	-2	4	2	4
23	26	-1	1	-1	1
120	135	0	10	0	10

$$\overline{X} = \frac{120}{5} = 24 \qquad \overline{Y} = \frac{135}{5} = 27$$

$$s = \sqrt{\frac{10 + 10}{5 + 5 - 2}} = \sqrt{\frac{20}{8}} = \sqrt{2.5}$$

$$s = 1.6 \qquad s^2 = 2.5$$

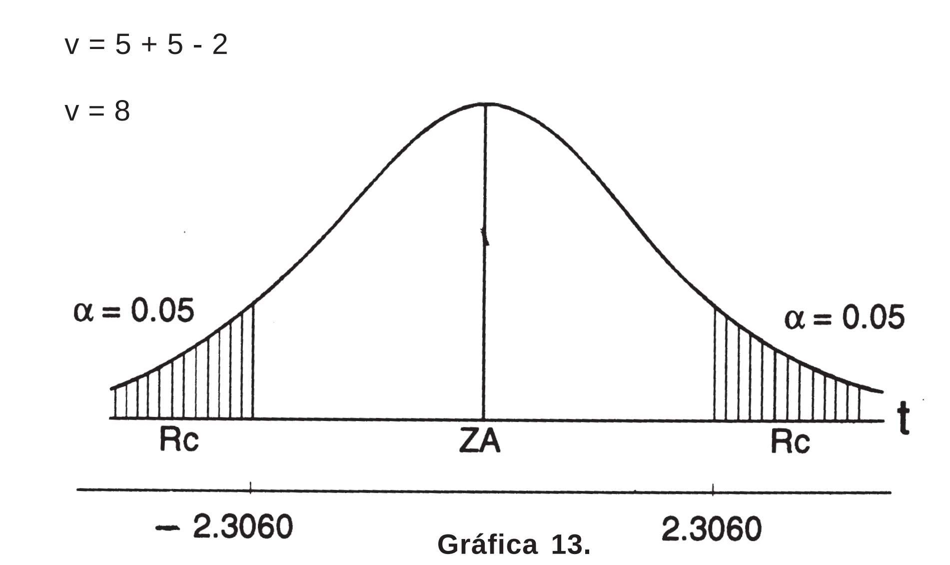
$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{2.5}{5} + \frac{2.5}{5}} = \sqrt{1} = 1$$

d) $t = \dfrac{(24 - 27) - 0}{1} = \dfrac{-3}{1} = -3$

e) $v = n_1 + n_2 - 2$

$v = 5 + 5 - 2$

$v = 8$



Gráfica 13.

f) Los resultados anteriores señalan que existe una diferencia significativa en el contenido medio de nicotina en ambas marcas. Véase gráfica 13.

5.2.3 *Distribución Chi Cuadrado:* χ^2

La distribución normal se utiliza en todos aquellos casos que ofrecen dos resultados posibles; cuando se presentan más de dos resultados posibles, debe aplicarse la prueba chi cuadrado que se simboliza así: χ^2. Un ejemplo típico de distribución lo constituye el lanzamiento de una moneda con posibilidades de que aparezca cara o sello. Uno de χ^2 consiste en el lanzamiento de un dado con seis caras posibles, numeradas del 1 al 6.

El Chi cuadrado es la suma de las fracciones que tienen por numerador el cuadrado de las diferencias entre las frecuencias reales u observadas y las frecuencias esperadas o teóricas y por denominador la frecuencia esperada.

$$\chi^2 = \sum \frac{(n_i - n^*_i)^2}{n^*_i}$$

en donde n_i frecuencia observada o real

n^*_i frecuencia teórica o esperada

Se puede observar en la fórmula, que mientras mayor sea la coincidencia entre las frecuencias observadas y las esperadas, menor será el valor de χ^2. Si $\chi^2 = 0$ significa concordancia entre las frecuencias observadas y las esperadas.

Ejemplo:

Supongamos que se lanza un dado 60 veces. Se sabe que las frecuencias teóricas, para este caso son de 10 veces cada cara y las frecuencias reales son los resultados del lanzamiento. ¿La base de un nivel de significancia del 5%, permite suponer que el dado no es perfecto?

Solución:

El problema da las frecuencias reales nt, para cada cara, como se ve en la tabla 12. Las frecuencias esperadas se obtienen multiplicando la probabilidad de cada suceso por n (número de lanzamientos). Véase gráfica 14.

Tabla 12. Frecuencia en el lanzamiento de un dado 60 veces

Caras	n_i	n_i^*	$n_i - n_i^*$	$(n_i - n_i^*)^2$	$\frac{(n_i - n^*_i)^2}{n^*_i}$
1	7	10	-3	9	0.9
2	14	10	4	16	1.6
3	8	10	-2	4	0.4
4	5	10	-5	25	2.5
5	16	10	6	36	3.6
6	10	10	0	0	0
Σ	60	60	0	-	9.0

$n^*_i = np_1 = 60\left(\frac{1}{6}\right) = \frac{60}{6} = 10$ y así se obtienen las demás frecuencias teóricas.

Se realiza la prueba teniendo en cuenta los siguientes pasos:

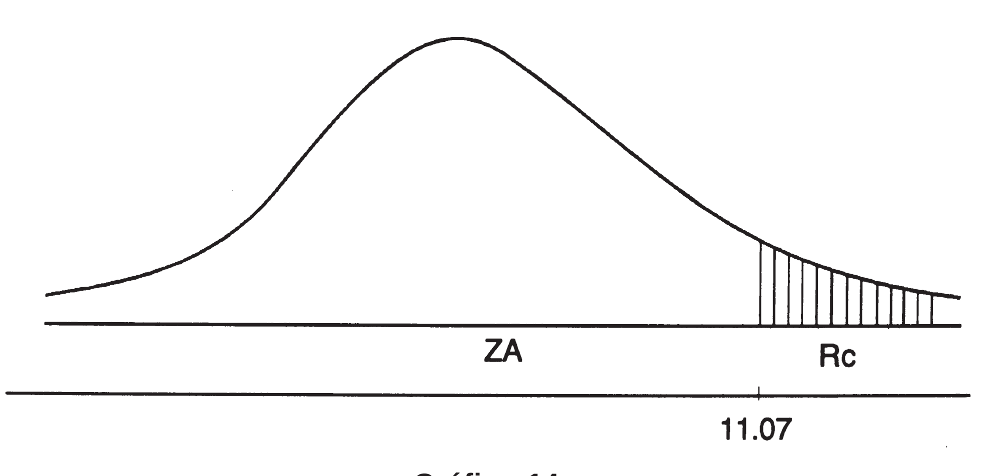
1) $H_I: n_i \neq n^*_i$

$H_I: n_i \neq n^*_i$

2) $\alpha = 0.05$

3) $\chi^2 = \sum \frac{(n_i - n^*_i)^2}{n^*_i} = 9.0$

4) $V = n - 1 = 6 - 1 = 5 \therefore \chi^2_{0.05} = 11.07$



Gráfica 14.

En la tabla de distribución de χ^2, encontramos que 11.07 es el valor crítico.

5) $\chi^2 < \chi^2_{0.05}$, es decir 9 < 11.07, por lo tanto se acepta la hipótesis de que la diferencia no es significativa. En otras palabras podemos afirmar que a un nivel del 5%, las diferencias que presentan las frecuencias reales, con relación a las frecuencias teóricas no nos da base para afirmar que el dado está cargado.

PRUEBAS DE HIPÓTESIS - AUTOEVALUACIÓN N° 5

1. ¿La hipótesis de investigación es exactamente lo mismo que la hipótesis nula?

   Sí _____ No _____

2. Si una hipótesis de investigación es que dos poblaciones tienen iguales medias aritméticas, ¿cuál es la hipótesis nula?

   a. Que en efecto tienen iguales medias aritméticas ____

   b. Que las dos poblaciones tienen distintas medias aritméticas ____

   c. Que las medias aritméticas pueden ser iguales o distintas ____

   d. Para poder definirla hay que conocer la desviación estándar ____

3. ¿Una distribución de muestreo es lo mismo que una distribución de una muestra?

   Sí _____ No _____

   ¿Por qué?
   ____
   ____

4. El nivel de significancia es:

   a. Lo mismo que el nivel de precisión _____

   b. La región crítica de una curva normal _____

   c. Una probabilidad _____

   d. El error alfa α _____

   e. La exactitud con que se predicen los parámetros _____

5. El valor crítico de un puntaje Z para una prueba de una cola, al nivel de significancia 0.05 es: ________

6. ¿A partir de una muestra de obreros en una fábrica, se puede inferir a cerca de las características de esa fábrica?

   a. No _____

   b. Sólo si uso una prueba Z _____

   c. Sólo si uso una prueba t _____

   d. Si se usa una prueba Z o prueba t _____

7. ¿La prueba t se prefiere a la prueba Z cuando no se conocen los parámetros del universo?

   Sí _____ No _____

8. Los salarios diarios de una industria están distribuidos normalmente con una media de $132 y una desviación estándar de $25. Si una empresa de dicha industria, que cuenta con 40 obreros paga en promedio $122, puede acusarse a esta compañía de pagar salarios inferiores al nivel de significancia del 1%?

   Sí _____ No _____

   Desarrollo: ____
   ____
   ____

9. Una hipótesis dice que el estudiante promedio de la universidad colombiana tiene un coeficiente de inteligencia mayor que el resto de la población. Escriba una hipótesis de investigación y una hipótesis nula, con un coeficiente de inteligencia que al estandarizarse tiene una $\overline{X} = 100$

   Hipótesis de investigación: ______________________________

   Hipótesis nula: ____

10. Un fabricante de ciertas piezas de proyectiles sostiene que en condiciones normales de reparación, tienen una duración media μ=320 horas. Probar esta afirmación frente a la alternativa $\mu \neq$320, si 16 piezas duran un promedio de 308 horas, con una desviación de 29 horas. Utilizar un nivel de significancia del 5%.

11. Un dado se lanza 100 veces y se obtienen los siguientes resultados:

Cara A	1	2	3	4	5	6
ni	12	17	20	22	13	16

¿Puede considerarse el dado perfecto al nivel del 1% de significancia?

Sí _____ No _____

Explique.

RESPUESTAS A LA AUTOEVALUACIÓN Nº 5

1. No, es lo contrario.
2. b
3. No. La distribución de una muestra son las características que tienen los datos a partir de una muestra.
   La distribución de un muestreo es la distribución de una muestra de medias a partir de muestras.
4. c
5. 1.96
6. d
7. Sí
8. $Z = \dfrac{122 - 132}{25/\sqrt{40}} = -2.53$ cae en la región de rechazo. Se puede acusar a la compañía de pagar salarios inferiores al nivel del 1%
9. Hipótesis de investigación: $H_1 = \bar{x} > 100$ Hipótesis nula: $H_0 = \bar{x} \leq 100$
10. $t = \dfrac{308 - 320}{29/\sqrt{15}} = -1.60$ cae en la zona de aceptación, o sea, que la duración promedio es de 320, a un nivel significativo del 5%.
11. Sí, porque $\chi^2 = 4.518$ y cae en la zona de aceptación y se puede considerar el dado perfecto.

6. REGRESIÓN Y CORRELACIÓN

6.1 Introducción a la bidimensional

En esta parte examinaremos la relación entre variables, medidas en escalas de intervalo o de razón. Pero ya no sólo queremos comparar, ahora queremos entender qué tipo de conexión hay entre las variables, por intermedio de la relación que existe entre ellas, e ir más allá entendiendo la naturaleza de la relación entre las variables.

En las distribuciones bidimensionales se consideran dos variables en forma simultánea, determinándose si tienen alguna relación funcional entre sí; aún más, cuantificando dicha relación. Estas variables pueden ser ambas discretas, ambas continuas, o una discreta y otra continua.

Llamaremos Xi a la primera variable, donde i toma todos los valores desde 1 hasta n. Así tendremos X_1, X_2, X_3, ... X_n; Y_i se considerará como la segunda variable, donde i toma los valores desde 1 hasta n. Se tendrán tantos valores de X como valores de Y, es decir, se tomarán siempre pares de observaciones.

Tabla. Valores de X y Y

X_i	Y_i	X_i	Y_i	X_i^2	Y_i^2	XY
X_1	Y_1	6	3	36	9	18
X_2	Y_2	10	4	100	16	40
X_3	Y_3	14	8	196	64	112
X_4	Y_4	20	12	400	144	240
		27	18	729	324	486
		33	25	1.089	625	825
Xn	Yn	110	70	2.550	1.182	1.721

Para cada variable se podrá calcular la media aritmética, la varianza y la desviación estándar, usando los datos no agrupados.

$$\overline{X} = \frac{\sum X_i}{N} = \frac{110}{6} = 18.33$$

$$\overline{Y} = \frac{\sum Y_i}{n} = \frac{70}{6} = 11.67$$

$$S_x^2 = \frac{\sum X_i^2}{n} - \overline{X}^2 = \frac{2.550}{5} - (18.33)^2 = 89.01$$

$$S_y^2 = \frac{\sum Y_i^2}{n} - \overline{Y}^2 = \frac{1.182}{6} - (11.66)^2 = 61.04$$

$$S_y = \sqrt{S_y^2} = \sqrt{89.01} = 9.44$$

$$S_y = \sqrt{S_y^2} = \sqrt{61.04} = 7.81$$

La covarianza es una medida de dispersión que representa el grado de variabilidad conjunta entre X_i y Y_i. Su símbolo es m_{xy} ó m_{yx}. Se define como la media del producto de las desviaciones respecto a las medias aritméticas. Entonces se puede calcular la covarianza por la siguiente fórmula:

$$m_{xy} = \frac{\sum XY}{n} - \overline{X} \cdot \overline{Y}$$

Tomando los datos de la tabla anterior podemos calcular la covarianza así:

$$m_{xy} = \frac{1.721}{6} = - (18.33)(11.67)$$

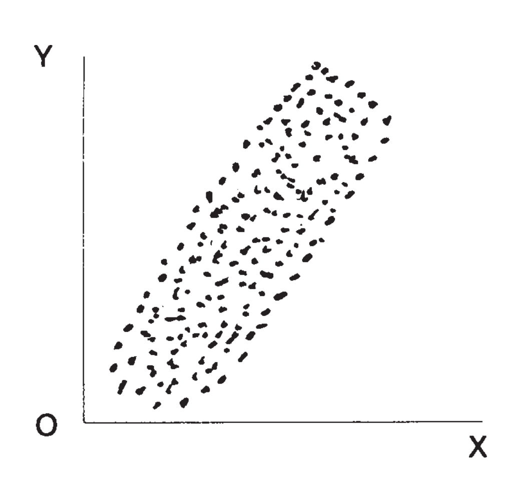
$$m_{xy} = 286.83 - 213.91$$

$$m_{xy} = 72.92$$

Los datos de la distribución bidimensional anteriormente dada puede representarse gráficamente en un par de ejes de coordenadas, tomando el eje de la abscisas para la primera variable (X_i) y el eje vertical para los valores de la segunda variable (Y_i).

En un plano cartesiano se presentan tantos datos como puntos pares de observaciones se tengan, correspondiendo cada punto a un par de observaciones. A esta representación gráfica se le denomina indistintamente diagrama de desparramiento, de esparcimiento o nube de puntos.

Gráfica 15. Nube de puntos



El problema consiste ahora en unir varios puntos de ese conjunto o nube de puntos, mediante un ajuste, ya sea rectilíneo, parabólico, exponencial o de cualquier otro tipo de línea que represente al conjunto. En algunos casos esos puntos estarán condensados al rededor de la línea, en otros presentarán diferencias; en este último caso, se pueden producir grandes errores, como consecuencia de que el ajuste realizado no es el más indicado.

El tipo de línea que se relaciona, dependerá de la forma que asuma el conjunto de puntos, al hacer la respectiva gráfica. También valiéndose de la experiencia del estadístico se puede determinar el mejor ajuste. En la exponencial, es muy sencilla la identificación, basta que la variable muestre un crecimiento geométrico, como por ejemplo, la población, el producto bruto, etc.

En general, se dice que la curva que hace mínima la suma de los cuadrados de las desviaciones entre puntos dados, dicha línea es la mejor.

6.2 Ajuste de una recta de regresión rectilíneo simple

Supóngase que se desea ajustar una recta, para ello sabemos que la ecuación general de la recta es:

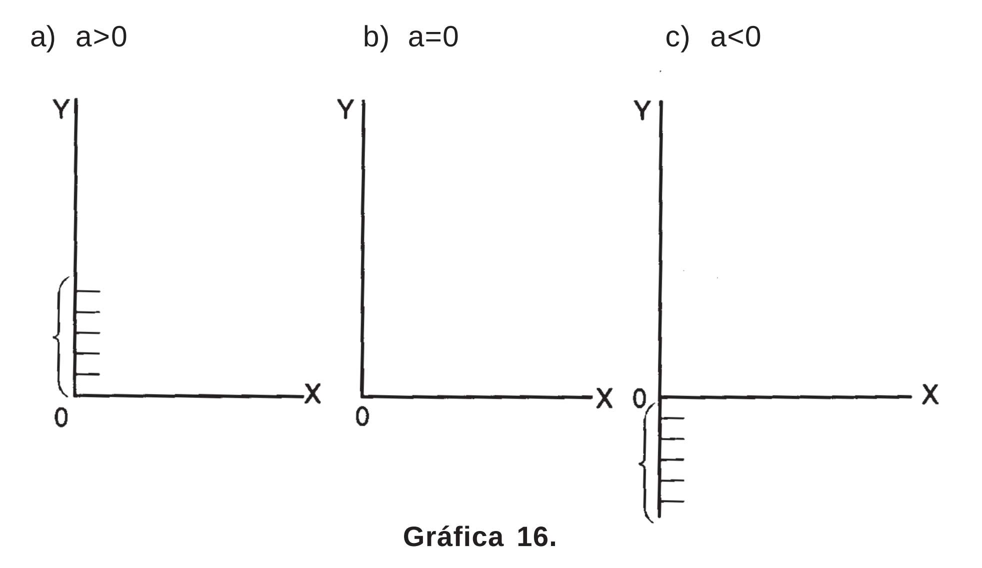
$Y = a + bx$, $X = a + by$

donde X, en la primera ecuación y Y en la segunda es la variable que se supone conocida, llamada variable independiente. Ahora, Y y X en una segunda ecuación corresponde a la variable que se va a estimar, conocida con el nombre de variable dependiente.

a. es el coeficiente de posición, denominado también origen, o sea, es la altura de la perpendicular levantada en el punto de origen. Véase gráfica 16.

El coeficiente de posición puede ser mayor, menor o igual a cero.

a) $a>0$ b) $a=0$ c) $a<0$

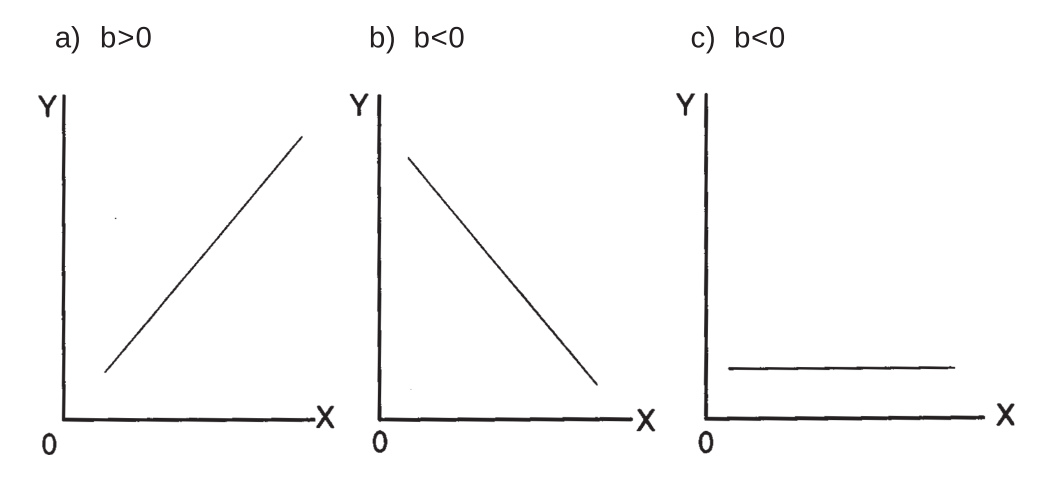


Gráfica 16.

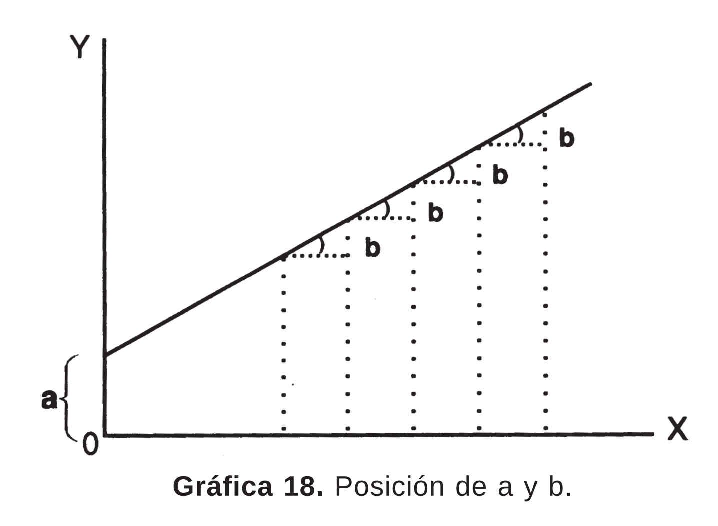
En el primer caso será un punto por encima del origen, en el segundo pasará por el origen y en el tercero estará por debajo del origen.

b. Es el coeficiente angular, el cual nos determina los crecimientos o la cantidad que aumenta en Y, por cada unidad que aumenta X y se denomina como pendiente.

El coeficiente angular puede ser:



Gráfica 17. Coeficiente angular b.



Gráfica 18. Posición de a y b.

Si b es un valor mayor que cero, es decir positivo, nos indicará que la recta es ascendente; si b es menor que cero, la recta será descendente y si es igual a cero, la recta es paralela al eje horizontal o vertical.

Cuando se tiene la ecuación Y = a + bx, se dice en este caso que se está estimando a Y en función de X. Para la ecuación X = a + by, decimos que estimamos a X en función de Y

El problema, ahora, consiste en calcular los parámetros a y b.

Partiendo de las ecuaciones normales de los mínimos cuadrados, se pueden calcular estos parámetros.

Las ecuaciones normales para los mínimos cuadrados son:

1) $\Sigma Y = Na + b\Sigma X$
2) $\Sigma XY = a\Sigma X + b\Sigma X^2$

Tomaremos los datos de la distribución bidimensional de la tabla y reemplazando en estas ecuaciones se tiene:

1) 70 = 6a + 110b
2) 1.721 = 110a + 2.550b

Por sustitución, eliminamos a y calculamos el valor de b así, multiplicamos los valores de la ecuación (1) por -110 y los de la ecuación (2) por 6 y se resta de la primera.

1) -7.700 = -660a - 12.100b

2) 10.326 = 660a + 15.300b

2.626 = 3.200b

$$b = \frac{2.626}{3.200} = 0.82$$

Reemplazamos el valor calculado de b en la ecuación (1) así:

1) 70 = 6a+110(0.82)

70 = 6a + 90.2

$$a = \frac{70 - 90.2}{6}$$

a = -3.37

Reemplazando en la ecuación general de la recta, Y = a + bx, se tiene:

Y = 3.37 + 0.82x

cuando se quiere estimar X en función de Y, se procede de la siguiente forma:

X = a + by

Las ecuaciones normales para el cálculo de los parámetros a y b son:

1) $\Sigma X = Na + b\Sigma Y$

2) $\Sigma XY.= a\Sigma Y + b\Sigma Y^2$

1) $110 = 6a + 70b$

2) $1.721 = 70a + 1.128b$

Multiplicando (1) por -70 y (2) por 6 y luego restamos de la primera:

1) $-7.700 = -420a - 4.900b$

2) $10.326 = 420a + 7.092b$

$2.626 = 2.192b$

$$b = \frac{2.626}{2.192} = 1.20$$

Reemplazando este valor en la ecuación (1) se tiene el valor de a

1) $110 = 6a+70(1.20)$

$110 = 6a+84$

$$a = \frac{100 - 84}{6}$$

$$a = \frac{100 - 84}{6}$$

Reemplazando en la ecuación general de la recta X = a + by, se tiene:

$X = 4.33 + 1.20y$

6.3 Correlación

En la práctica algunas de las variables que medimos en los procesos de investigación, tienen un nivel de medición intervalo o de razón y para resolver problemas con estos niveles existen unas técnicas diferentes a las que hasta ahora hemos visto.

Cuando tenemos mediciones realizadas con escalas intervalos o de razón, la mejor medida paramétrica de asociación es la relación de Pearson (r), también conocida como la relación Producto-Momento. Para hacer una correcta interpretación del coeficiente, se tienen que hacer dos suposiciones.

1. Se supone que la relación es lineal.

2. Que la distribución de las variables asociadas tiende a ser normal.

Es importante entender que una correlación entre dos variables produce un número que resume la naturaleza de la asociación entre las variables.

En el caso de la correlación de Pearson, ese número (r) toma valores desde -1 hasta +1, pasando por cero, cuando hay relación entre las variables. Si la relación tiende a ser inversa, se aproximará a -1, en la medida en que sea perfecta y por el otro lado, si la correlación tiende a ser directa, se acercará a +1. No sólo sirve para indicar si hay o no relación, sino para indicar la fuerza de la misma y también para compararla con las otras variables.

Una forma adecuada para ilustrar la naturaleza de r, se muestra en la figura 19. Aquí se presentan 6 tipos de relaciones entre las dos variables X y Y. Cada punto en los cuadrantes representa los valores de X, y Y toma un elemento de la muestra o población analizada.

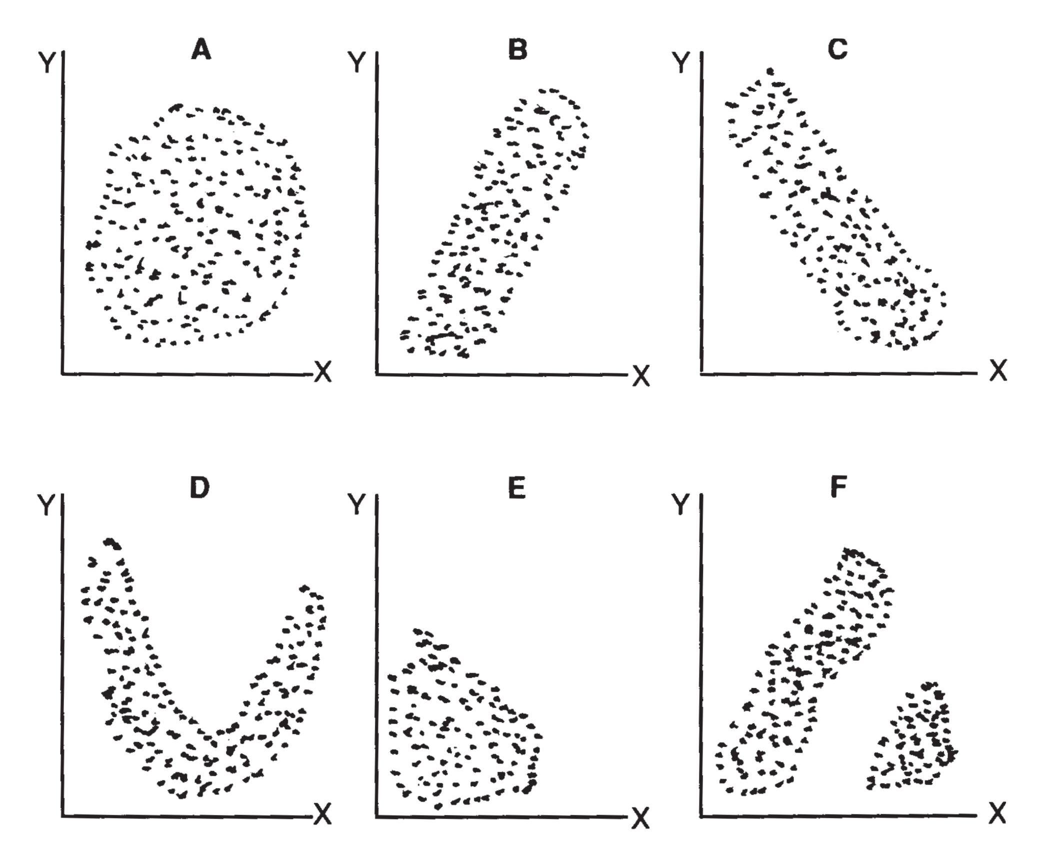
En la gráfica A, se ve claramente una nube de puntos dispersos, casi que aleatoriamente distribuidos. Cuando esto sucede no existe ninguna relación directa entre las dos variables, es decir r es igual o tiende a cero ($r \cong 0$).

En la gráfica B, vemos que existe una tendencia. A mayores valores de X parecen estar asociados mayores valores de Y, por lo tanto, existe una relación directa entre las dos variables. Es decir $r>0$.

En la gráfica C, vemos que también hay una tendencia, pero en este caso indicando que a valores pequeños de X corresponden valores grandes de Y y a valores grandes de X corresponden valores pequeños de Y. O sea, hay una relación inversa entre las dos variables $r<0$.

En la gráfica D, vemos que hay una correlación no lineal entre las dos variables. En el rango de valores bajos y altos de X corresponden valores bajos de Y y en el rango de valores intermedios de X corresponden valores altos de Y. Aunque $r \cong 0$, no podemos decir que no hay relación entre las

dos variables, sólo que la relación es curvilínea y por lo tanto requiere de un tratamiento particular.



Gráfica 19. Ejemplos para entender la interpretación de r.

Los dos últimos gráficos E y F sirven para ilustrar una situación que suele presentarse durante una investigación.

En el caso E, vemos que la mayoría de los elementos analizados, parecen indicar la no existencia de relación entre las dos variables, pero hay unos pocos casos donde altos valores de X tienen a los valores de Y.

En el caso F, se presenta lo contrario, la mayoría de las observaciones analizadas indican una relación positiva entre X y Y, pero unos pocos casos la distorsionan, ya que rompen la tendencia.

Estos ejemplos, muestran la utilidad de hacer uso de histogramas y de las distribuciones de las frecuencias obtenidas, para no cometer errores. Por ejemplo, concluir que no hay relación en el caso del gráfico D, cuando

en realidad hay una no lineal; o decir que hay relación positiva en el gráfico E, o que no existe relación en el gráfico D, cuando en realidad es lo contrario.

En estos dos últimos casos, lo recomendable es retirar los elementos distorsionadores, o completar una mayor información, por si en realidad lo sugerido por los casos extremos es la tendencia real, o por último, revisar los procesos de recolección, codificación o sistematización de la información, donde se pudo haber cometido algún error.

6.4 Coeficiente de correlación

Nos permite cuantificar el grado de correlación existente entre las variables X y Y.

Hay dos maneras de calcular el coeficiente de correlación, una aritmética y otra por medio de álgebra de matrices.

Hagamos un ejercicio aritmético:

Supongamos que queremos medir la asociación que existe entre innovación tecnológica y producción, en 11 empresas que elaboraron un mismo producto. En la tabla 12 se presenta la información obtenida en visitas realizadas a 11 industrias.

Tabla 12. Ejemplo para calcular r

Empresa	Nivel tecnológico X	Producción y	X^2	Y^2	XY
1	44	20	1.936	400	880
2	62	19	3.844	361	1.178
3	30	15	900	225	450
4	50	21	2.500	441	1.050
5	28	13	784	169	364
6	34	15	1.156	225	510
7	49	17	2.401	289	833
8	54	17	2.916	289	918
9	72	22	5.184	484	1.584
10	66	21	4.356	441	1.386
11	27	14	729	196	378
N=11	X=516	Y=194	X^2= 26.706	Y^2=3.520	XY = 9.531

Una fórmula para calcular r es:

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[N\Sigma X^2 - (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

Sustituyendo tenemos:

$$r = \frac{(11)(9.531) - (516)(194)}{\sqrt{\left[11(26.706) - (516)^2\right]\left[11(3.520) - (194)^2\right]}}$$

$$r = \frac{4.737}{\sqrt{29.820.840}}$$

r = 0,867

Como correlaciones cercanas a 1 son altas e indican una asociación positiva, concluimos que hay una alta asociación entre la tecnología utilizada y la producción en el tipo de industria estudiadas.

El grado de correlación lo podemos clasificar evitando un tanto la rigidez de sus límites:

a. Correlación perfecta, cuando r=1 (o menos de 1).

b. Correlación excelente, cuando r es mayor que 0.90 y menor que 1, ó (-1<Y<-0.90).

c. Correlación aceptable, cuando r se encuentra entre 0.80 y 0.90 ó (-090<r<-080).

d. Correlación regular, cuando r se encuentra entre 0.60 y 0.80 (-080<r<-0.60).

e. Correlación mínima, cuando r se encuentra entre 0.30 y 0.60 (-0.60<r<-0.30).

f. No hay correlación para r menor de 0.30 ó (-0.30<r<0)

REGRESIÓN Y CORRELACIÓN - AUTOEVALUACIÓN N° 6

1. La fórmula Y = a + bx, ¿corresponde a la ecuación general de la recta?

   Sí _____ No _____

   ¿Qué nombre recibe la variable X? _________________________

2. ¿Qué nombre recibe a?

   ____

3. ¿Qué nombre recibe b?

   ____

4. Con base en los datos de la siguiente tabla de valores para X y Y, halle la recta de regresión por los mínimos cuadrados. Para Y=a+bx

Xi	Yi
13	5
14	7
8	9
10	8
15	11
60	40

5. Calcule el coeficiente de correlación, a partir de los datos de la siguiente tabla:

Xi	Yi
9	23
17	35
20	29
19	33
20	43
23	32
108	195

RESPUESTAS A LA AUTOEVALUACIÓN N° 6

1. Sí. Variable independiente.
2. Coeficiente de posición.
3. Coeficiente angular.
4. $Y = -180.4 + 15.7x$
5. $r = 0.60$

7. ANÁLISIS DE LA VARIANZA

Del análisis de varianza, podemos decir que esta técnica estadística, normalmente es utilizada para analizar resultados en la investigación con diseños experimentales y cuasi experimentales. Muchas veces necesitamos comparar dos o más distribuciones, que corresponden a variaciones de una misma variable dependiente, afectada por una o más variables independientes.

Las variables independientes pueden ser medidas en cualquier tipo de escala, pero la variable dependiente debe ser medida, al menos, al nivel intervalo. Si las variables independientes son medidas al nivel intervalo o de razón, tendríamos un típico caso de análisis de regresión múltiple.

Si queremos medir la importancia relativa de tres procesos en la calidad de un producto final utilizamos el análisis de varianza (ANOVA).

En la tabla 13 tomamos unos datos hipotéticos para ilustrar su caso. Tenemos tres grupos de 5 productos, seleccionados independientemente, al azar, con sus respectivos puntajes, en cuanto a su calidad; ésta, que es la variable dependiente, se midió por medio de una escala intervalo.

Tabla 13. Datos para ilustrar el uso de Anova

Método 1		**Método2**		**Método 3**		**Total**
X	**X^2**	**X**	**X^2**	**X**	**X^2**	
6 10 9 11 12	36 100 81 121 144	9 7 8 8 5	81 49 64 64 25	4 6 9 4 3	16 36 81 16 9	
Σ 48	482	37	283	26	158	111
$\bar{x}$ 9.6		7.4		5.2		7.4
N 5		5		5		1.5

Supongamos que cada grupo sometido a un método de producción distinto, tiene varianzas iguales. Es decir, la hipótesis nula H_0, es que no

hay diferencias en las medias aritméticas de los tres grupos. Supongamos que queremos analizar las observaciones a un nivel de significancia del 0.05 y ver si hay diferencias en la calidad por el método empleado. El cálculo de F, contra cuya distribución de muestreo se comparan los resultados, es el procedimiento a seguir.

1° Suma total de los cuadrados de los puntajes de los tres grupos así:

$$\sum(X-\overline{X})^2 = \sum X^2 - \frac{(\sum X)^2}{N}$$

$$= (482 + 283 + 158) - \frac{(111)^2}{15}$$

$$= 923 - 821.4 = 101.6$$

2° Suma de cuadrados entre columnas. Igual fórmula pero referida a cada uno de los tres grupos o métodos utilizados así:

$$\text{Método 1} = 482 - \frac{(48)^2}{5} = 482 - 460.8 = 21.2$$

$$\text{Método 2} = 283 - \frac{(37)^2}{5} = 283 - 273.8 = 9.2$$

$$\text{Método 3} = 158 - \frac{(26)^2}{5} = 158 - 135.2 = \frac{22.8}{53.2}$$

Total

3° Suma de cuadrados entre columnas (grupos).

Suma total del cuadrado = suma “dentro” + suma “entre”

Suma “entre” = suma total - suma “dentro”.

Suma “entre” = 101.6 - 53.2

Suma “entre” = 48.4

4° Grados de libertad. Son los siguientes:

Suma total = n-1

= 15-1=14

Suma "dentro" = $K(N_1-1)$; donde K=número de grupos y N_1 tamaño de los grupos (5)

= 3(5-1)

= 12

Suma "entre" = K-1

= 3-1

= 2

5° Cálculo para análisis de la varianza con el valor de F

	Suma de Cuadrados	**Grados de libertad**	**Estimación de la varianza**	**F**
Total	101.6	14		
Entre columnas	48.4	2	24.20	
Dentro de las columnas	53.2	12	4.43	5.46

La estimación de la varianza es la razón de la suma de cuadrados y los grados de libertad y F es igual a la razón de la estimación "entre" y la estimación "dentro".

$$\text{Varianza "entre"} = \frac{48.4}{2} = 24.2$$

$$F = \frac{24.2}{4.43} = 5.46$$

$$\text{Varianza "dentro"} = \frac{53.2}{12} = 4.43$$

6° Decisión

En la tabla de distribución F vemos que con un nivel de significancia del 0.05 y con 2 y 12 grados de libertad, necesitamos un valor igual o mayor a 3.88 para poder rechazar H_0. Como en efecto obtuvimos uno de 5.46, rechazamos H_0 y concluimos que los métodos utilizados sí producen una diferencia significativa en la calidad del producto.

ANÁLISIS DE VARIANZA - AUTOEVALUACIÓN N° 7

1. En una universidad se quiere probar el efecto que distintos métodos de enseñanza tienen, en los resultados de un curso dado. Tres métodos se proponen y se aplican a tres grupos, compuestos de 14 estudiantes, cada grupo seleccionado aleatoriamente entre los alumnos del curso. Los resultados del examen final del semestre se presentan a continuación. Se puede concluir que los métodos producen un efecto diferencia.

52	27	16
49	36	14
34	34	23
52	32	21
14	41	29
41	33	20
34	25	28
42	19	27
33	28	17
39	22	12
51	25	40
43	14	28
39	22	18
43	27	21

2. La distribución que se utiliza para probar el nivel de significancia de un análisis de varianza es el χ^2

Verdadero _____ Falso _____

1. Suma total = 5.112.4

2. Suma de cuadrados entre columnas:

   Método 1 = 1.309.4

   Método 2 = 675.5

   Método 3 = 715.4

   Total 2.700.3

3. Suma de cuadrados entre columnas.

   Suma de cuadrados entre las columnas:

   5.112.4

   2.700.3

   2.412.1

4. Grados de libertad.

   Suma total = 42-1 =41

   Suma dentro = 3(14 - 1)39

   Suma entre = 3-1 =2

5. Cálculo F

	Sumade cuadrados	Grados de libertad	Varianza	F
Total	5.112.40	41		
Entre	2.412.10	2	1.206.10	
Dentro	2.700.30	39	69.24	17.42

RESPUESTA A LA AUTOEVALUACIÓN N° 7

Si se prueba con significancia del 0.01, esto implica que necesitamos un F al menos de 5.3, para rechazar H_0 y aceptar que los métodos sí tienen efecto en el rendimiento académico de los estudiantes y como se obtiene un F=17.42, concluimos que sí tiene efectos.

2. Falso, es la distribución F.

Comente los resultados de los 2 problemas.

8. ESTUDIO DE FACTIBILIDAD DEL SOFTWARE ESTADÍSTICO

La necesidad de contar con una herramienta adecuada para la enseñanza e investigación estadística, conlleva a realizar un estudio detallado de los programas con mayor aplicación en estos campos, además se debe considerar para la implementación de éstos, los recursos físicos para un desempeño eficaz.

El paquete estadístico SAS y SPSS cumplen con los requerimientos académicos debido a su manejo personalizado, cursos de actualización y asesoría continua, lo cual contribuye como soporte a la docencia y en gran parte al proceso educativo.

8.1 SAS

Es un sistema que provee herramientas para la captura, administración y análisis de datos.

- Este software permite un total control sobre la información.
- Permite accesar, administrar, analizar y presentar la información.
- El SAS está conformado por los siguientes módulos:

STAT

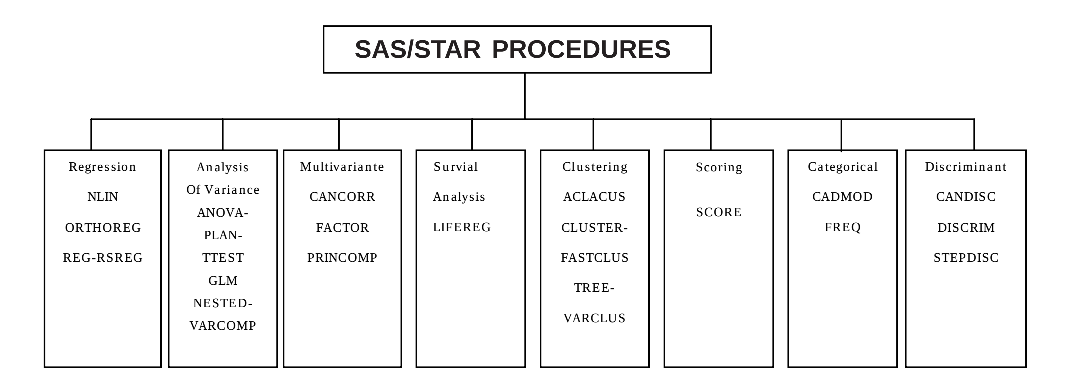
Provee herramientas para el análisis estadístico avanzado. Sirve para administrar la información como:

– Almacenamiento y recuperación de información.

– Programación y modificación de datos.

– Reportes.

– Análisis estadístico.



FSP**

Sirve para construir y manipular conjuntos de datos. Este módulo facilita la captura de datos, validación, verificación en tablas, edición y consulta.

AF**

Sirve para crear menús y para encadenar los diversos módulos.

ETS*

Provee herramientas para realizar proyecciones en modelos econométricos. Puede realizar reportes como tablas de amortización y tablas de depreciación.

OR

Provee una herramienta para investigación de operaciones.

QC*

Es una aplicación para estadísticas en control de calidad.

GRAPH

Es el módulo que posee el **SAS**

Convenciones:
* No es necesario.
** Para realizar aplicaciones bajo ambientes SAS.

SHARE

Permite el acceso concurrente de varios usuarios a archivos comunes **SAS.** Este módulo garantiza la integridad y la seguridad de la información.

ACCES

Este módulo sirve para realizar las interfaces en los manejadores de bases de datos.

8.2 SPSS

BASE

Contiene estadísticas básicas, gráficos de alta resolución y un paquete completo de listado. Con el **SPSS BASE 6.0** bajo Windows se producirá un análisis y resultados de alta calidad.

– **Estadísticas profesionales**

Medidas de similariedad y diferencias en sus datos, provee técnicas de clasificación. En este procedimiento incluye Análisis Cluster, Análisis Discriminante, Análisis Factor, Análisis Multidimensional, Análisis Proximidades, Mínimos Cuadrados con Ponderación y Mínimos Cuadrados con dos fases.

– **Estadísticas avanzadas**

Incluye sofisticadas técnicas como: Regresión Decox, Estimación Kalpian Meier, Regresión Logística, Análisis Lineal Logarítmico, Análisis de Varianza Multivariado, Regresión No-Lineal de Vida (Tabla vida).

– **Tablas**

Crea una variedad de cuadros tabulares, incluyendo tablas complejas (múltiple respuesta).

– **Tendencias**

Ejecuta pronósticos y análisis de series de tiempo, modelos de suavización y métodos para estimación autorregresión.

– **Categorías**

Ejecuta análisis conjunto y procedimientos de optimización, incluyendo análisis de correspondencia.

– **Chaid**

Desarrolla modelos predictivos, produce variables predictorias.

– **Lissrel**

Relaciona un análisis estructural lineal y modelos de ecuaciones simultáneas.

8.3 Requerimientos mínimos para la instalación del Software SAS

1. Procesador 486 o superior.

2. Memoria RAM 32 Mb o superior.

3. 20 Mb de espacio en el disco duro.

4. Cualquiera de las siguientes opciones

   • Windows versión 3.1 o posteriores.
   • Windows para trabajo en grupo versión 3.1 o posteriores.

8.4 Requerimientos mínimos para la instalación del Software SPSS

1. Procesador 386 o superior.

2. Memoria RAM 32 Mb o superior.

3. 20 Mb como mínimo para almacenar el SPSS.

4. Monitor VGA.

5. Cualquiera de las siguientes opciones.

6. Windows versión 3.1 o posteriores.

   • Windows para trabajo en grupo versión 3.1 o posteriores.

8.5 Manejador de Base de Datos

En la actualidad existen en el mercado una gran variedad de manejadores de bases de datos con características únicas, es decir, no son compatibles con los sistemas tradicionales de manejadores de bases de datos de segunda y tercera generación como Dbase, DataEase, Foxpro. Los manejadores de Bases de Datos de cuarta generación que se recomiendan por compatibilidad son Acces, Paradox, Foxpro.

Para cumplir con los objetivos educativos se recomienda por su flexibilidad y compatibilidad el Foxplus bajo ambiente Windows.

Visual Foxpro 6.0

La herramienta para el desarrollador profesional de aplicaciones xBase. (En esta nueva versión obtendrá: Mejor rendimiento e IDE, mayor conectividad y soporte de ActiveX).

Características

Sigue la evolución de Foxpro desde un sistema de desarrollo de bases de datos de escritorio y de tipo procedural a un entorno de desarrollo orientado a objetos, con las herramientas necesarias para construir aplicaciones y componentes en sistemas clientelservidor e Internet.

Visual Foxpro 6.0 es un miembro de la familia que integra el sistema de desarrollo Visual Studio. Permite la utilización de los últimos avances en los sistemas operativos Windows y Windows NT y ha sido diseñado para permitir a los 500.000 desarrolladores que usan Visual Foxpro una forma más potente y eficiente de crear aplicaciones multicapa cliente/servidor, basadas en Web y centradas en el tratamiento de datos. La compatibilidad con el año 2000 está totalmente asegurada.

Visual Foxpro mantiene un completo set de comandos Xbase para permitir una curva suave de aprendizaje a los programadores de bases de datos en entornos MsDos.

Dispone de las características más avanzadas en el diseño de clases orientadas a objeto incluyendo herencia, subclases, encapsulación y polimorfismo. Las librerías de clases visuales y no visuales (por código) reducen enormente el tiempo de desarrollo.

Existen herramientas de diseño para todas las fases de desarrollo de la aplicación. Un motor de bases de datos altamente eficiente, un lenguaje centrado en los datos y la capacidad de creación de componentes hacen de Visual Fospro una herramienta idónea para la generación de lógica de negocio en los entornos multicapa con tratamientos intensivos de datos.

Visual Foxpro puede intercambiar datos con bases de datos SQL a través de ODBC (Open Database Connectivity). De esta forma, no es necesario un gran esfuerzo en la adaptación de aplicaciones basadas en servidor de ficheros a aplicaciones clientelservidor. También podemos utilizar los componentes ADO (ActiveX Data Objects) para intercambiar información con datos relacionases y no-relacionales mediante OLE DB.

Visual Foxpro puede ser utilizado en las tres capas de una arquitectura clientelservidor. Puede suministrar el interface de usuario a través del uso de formularios con toda la potencia de la orientación a objeto. Los controles ActiveX pueden usarse en los formularios y se subciasean para extender sus funcionalidades. La lógica de negocio puede encapsularse en componentes muy eficientes gracias a la potencia de Fospro en la recuperación y manipulación de datos. Los componentes COM pueden ser llamados desde el front-end del usuario o desde el servidor Internet. Los componentes comunican con las bases de datos de Fospro y SQL a través de ODBC y OLE DB. En almacenamiento de datos puede estar soportado por el motor de Visual Foxpro. En arquitecturas CIS la mejor combinación es SQL para el proceso de las transacciones y Visual Foxpro para el manejo de consultas locales y procesos batch.

Novedades en Visual Foxpro 6.0

La aparición de Visual Foxpro 6.0 ha supuesto principalmente una evolución de las herramientas de desarrollo para adaptarlas a la arquitectura COM y al desarrollo de aplicaciones multicapa en Internet e Intranet. Además dispone de herramientas de desarrollo mejoradas y ampliadas. Por último se han revisado los comandos de programación, destacando la mejora en el soporte del año 2000.

Las principales novedades son:

OLE Drag and Drop

El OLE Drag and Drop permite intercambiar datos entre los controles de una aplicación o entre varias aplicaciones que soporten esta funcionalidad. Ahora podemos arrastrar ficheros desde el explorador de Windows a la ventana del proyecto o transportar textos con el ratón desde Word o Excel.

Visual foxpro disponía desde versiones anteriores de un drag and drop propietario. Sigue existiendo, pero se recomienda no utilizar los dos tipos de drag and drop simultáneamente en una aplicación.

Documentos activos

Visual Foxpro permite crear documentos activos, los cuales permiten visualizar documentos no-Htmi en un explorador de Internet. De momento sólo es posible con el Microsoft Internet Explorer. Un documento activo es un tipo de documento OLE embebido. Se visualiza totalmente en un área de la aplicación contenedora o host, mezclando su menú con el del host. La tecnología de documento activo permite visualizar múltiples tipos de documentos dentro de un único contenedor.

A diferencia de Visual Basic, los formularios de Visual Foxpro no requieren modificaciones para crear documentos activos. Es suficiente con iniciar la aplicación desde una clase basada en la clase base ActiveDoc. La clase base ActiveDoc aporta las propiedades, eventos y métodos para un documento activo y de esta forma interactuar con el host.

Mejoras en el servidor de automatización

Un servidor de automatización es un componente que expone su funcionalidad, la cual puede ser usada por otras aplicaciones a través de la automatización. Visual Foxpro puede crear servidores de automatización dentro o fuera del proceso.

Las mejoras en este tema se refieren al soport del Apartement Model Threading. De esta forma Visual Foxpro tiene soporte para Microsoft Transaction Server. Los componentes construidos con Visual Foxpro pueden ser manejados por el explorador de Microsoft Transacción Server y participar en las transacciones con otros componentes.

Otras mejoras de adaptación de los servidores de automatización a la arquitectura COM son: gestión optimizada del Runtime, soporte mejorado de las librerías de tipos, manejo de excepciones, paso de arrays, nuevas propiedades y métodos para la automatización...

Galería de componentes

La galería de componentes es una nueva herramienta que ayuda a agrupar y organizar objetos como las clases, librerías, formularios, botones. Los distintos objetos pueden organizarse en otros objetos como proyectos,

aplicaciones u otras agrupaciones diferentes. Estas agrupaciones visuales se pueden adaptar dinámicamente a las necesidades del desarrollador.

La galería de componentes incluye también las Visual Foxpro Foundation Classes. Son un conjunto de clases para mejorar las aplicaciones, reutilizando código probado y optimizado.

Estas clases pueden incluirse en nuestros proyectos directamente o crear subciases para adaptarlas a las distintas aplicaciones.

Aplicación de cobertura y control de rendimiento de código

La aplicación de cobertura genera información de las líneas que fueron ejecutadas en un determinado fichero. Por otro lado se puede configurar el control que se realiza sobre las líneas de código ejecutadas, cuántas veces se ejecutan, duración y muchos otros aspectos.

Acceso por programa al proyecto

En versiones anteriores de Visual Foxpro, el único acceso a los proyectos se realizaba de forma directa, abriendo la tabla pix a través del gestor de proyectos. Ahora se ha implementado un objeto proyecto para poder manipularlo por programa. El objeto proyecto actúa de intermedio entre un proyecto abierto y el desarrollador que puede interactuar directamente con el proyecto.

9. STATGRAPHICS

Software para el análisis de datos cuantitativos y cualitativos cuya estructura está sobre la base de menús. Existen versiones para trabajar bajo DOS y bajo Windows.

El menú principal de Statgraphics permite entrar a cualquiera de las 22 secciones divididas en seis grupos.

9.1 Grupo I: Data handling and system utilities

A. DATA MANAGEMENT. Menú para entrada de datos, definición de variables, importación y exportación de datos.

B. SYSTEM ENVIROMANT. Acceso a requerimientos y configuración del sistema.

C. REPORT WRITER AND GRAPHICS REPLAY. Configuración de impresoras.

D. PLOTTER INTERFACE. Interfase para plotter.

9.2 Grupo II: Plotting and descriptiva statistics

E. PLOTTING FUNCTIONS. Elaboración de gráficas estadísticas en dos y tres dimensiones.

F. DESCRIPTIVE METHODS. Resume datos, realiza distribuciones de frecuencia, obtiene estadísticas básicas.

G. ESTIMATION AND TEST. Elabora intervalos de confianza y pruebas de hipótesis para una o dos muestras. Permite hacer pruebas de bondad de ajuste.

H. DISTRIBUTION FUNCTIONS. Genera números aleatorios. Permite hacer cálculos con las principales funciones de probabilidad continuas y discretas.

I. EXPLORATOR Y DATA ANALYSIS. Realiza el análisis exploratorio de datos, diagrama de tallos y hojas, diagrama de caja, etc.

9.3 Grupo III: Anova and regression analysis

J. ANALYSIS OF VARIANCE. Realiza análisis de varianza de una o dos vías.

K. REGRESION ANALYSIS. Ajusta modelos lineales y no lineales en una y varias variables.

9.4 Grupo IV: Time series procedures

L. FORECASTING. Pronosticación con modelos de series temporales.

M. SMOOTHING. Procedimientos de promedios móviles

N. QUALITY CONTROL. Procedimientos de control de calidad estadísticos.

O. TIME SERIES ANALYSIS. Procedimientos para ajustar modelos ARIMA a series temporales.

9.5 Grupo V: Advanced procedures

P. CATEGORICAL DATA ANALYSIS. Tablas de contingencia, correspondencias simples, etc.

Q. MULTIVARIATE METHODS. Componentes principales, análisis de conglomerados, tests multivariados, etc.

R. NONPARAMETRIC METHODS. Pruebas no paramétricas con una y varias muestras.

S. SAMPLING. Determina tamaño de la muestra.

T. EXPERIMENTAL DESIGN. Métodos de diseño de experimentos, anova, manova, diseños 2^2, 2^3, etc.

9.6 Grupo VI: Mathematical procedures

U. NUMERICAL ANALYSIS. Métodos de análisis numéricos, raíces de funciones, método simple, álgebra, etc.

V. MATHEMATICAL FUNCTIONS. Glosario de funciones matemáticas, aplicaciones.

9.7 Requerimientos físicos para la instalación de Statgraphics

1. Sistema operativo DOS (Windows 3.1 o superior si se va a instalar la versión bajo Windows).

2. Procesador 386 o superior

3. El monitor se configura en la instalación.

4. 5 Mb para almacenar el STATGRAPHICS.

Tablas

Áreas bajo la curva normal

Fracción del área total (10.000) bajo la curva normal, correspondiente a distancias entre la media y las ordenadas situadas a z unidades de desviación estándar de la media.

Z	.00	.01	.02	.03	04	.05	.06	07	.08	0.9
0.0	0000	0040	0089	0120	0159	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2518	2549
0.7	2580	2612	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3718	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4083	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4430	4441
1.6	4452	4463	4474	4485	4495	4505	4515	4523	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4758	4762	4767
2.0	4773	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4865	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4862	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4980	4980	4981
2.9	4981	4982	4983	4984	4984	4984	4985	4985	4986	4986
3.0	4986.5	4987	4987	4988	4988	4988	4989	4989	4989	4990
3.1	4990.0	4991	4991	4991	4992	4992	4992	4992	4993	4993
3.2	4993.129									
3.3	4995.166									
3.4	4996.631									
3.5	4997.674									
3.6	4998.409									
3.7	4998.922									
3.8	4999.277									
3.9	4999.519									
4.0	4999.683									
4.5	4999.966									
5.0	4999.997133									

Fuente: Harold O. Rugg, Statistical Methods Applied to Education, Houghton Mifflin Company, Boston, 1917, apéndice al anexo III, pp. 389-390, con autorización del editor.

Números aleatorios

10 09 73 25 33	76 52 01 35 86	34 67 35 48 76	80 95 90 91 17	39 29 27 49 45
37 54 20 48 05	64 89 47 42 96	24 80 52 40 37	20 63 61 04 02	00 82 29 16 65
08 42 26 89 53	19 64 50 93 03	23 20 90 25 60	15 95 33 47 64	35 08 03 36 06
99 01 90 25 29	09 37 67 07 15	38 31 13 11 65	88 67 67 43 97	04 43 62 76 59
12 80 79 99 70	80 15 73 61 47	64 03 23 66 53	98 95 11 68 77	12 17 17 68 33
66 06 57 47 17	34 07 27 68 50	36 69 73 61 70	65 81 33 98 85	11 19 92 91 70
31 06 01 08 05	45 57 18 24 06	35 30 34 26 14	86 79 90 74 39	23 40 30 97 32
85 26 97 76 02	02 05 16 56 92	68 66 57 48 18	73 05 38 52 47	18 62 38 85 79
63 57 33 21 35	05 32 54 70 48	90 55 35 75 48	28 46 82 87 09	83 49 12 56 24
73 79 64 57 53	03 52 96 47 78	35 80 83 42 82	60 93 52 03 44	35 27 38 84 35
98 52 01 77 67	14 90 56 86 07	22 10 94 05 58	60 97 09 34 33	50 50 07 39 98
11 80 50 54 31	39 80 82 77 32	50 72 56 82 48	29 40 52 42 01	52 77 56 78 51
83 45 29 96 34	06 28 89 80 83	13 74 67 00 78	18 47 54 06 10	68 71 17 78 17
88 68 54 02 00	86 50 75 84 01	36 76 66 79 51	90 36 47 64 93	29 60 91 10 62
99 59 46 73 48	87 51 76 49 69	91 82 60 89 28	93 78 56 13 68	23 47 83 41 13
65 48 11 76 74	17 46 85 09 50	58 04 77 69 74	73 03 95 71 86	40 21 81 65 44
80 12 43 56 35	17 72 70 80 15	45 31 82 23 74	21 11 57 82 53	14 38 55 37 63
74 35 09 98 17	77 40 27 72 14	43 23 60 02 10	45 52 16 42 37	96 28 60 26 55
69 91 62 68 03	66 25 22 91 48	36 93 68 72 03	76 62 11 39 90	94 40 05 64 18
09 89 32 05 05	14 22 56 85 14	46 42 75 67 88	96 29 77 88 22	54 38 21 45 98
91 49 91 45 23	68 47 92 76 86	46 16 28 35 54	94 75 08 99 23	37 08 92 00 48
80 33 69 45 98	26 94 03 68 58	70 29 73 41 35	53 14 03 33 40	42 05 08 23 41
44 10 48 19 49	85 15 74 79 54	32 97 92 65 75	57 60 04 08 81	22 22 20 64 13
12 55 07 37 42	11 10 00 20 40	12 86 07 46 97	96 64 48 94 39	28 70 72 58 15
63 60 64 93 29	16 50 53 44 84	40 21 95 25 63	43 65 17 70 82	07 20 73 17 90
61 19 69 04 46	26 45 74 77 74	51 92 43 37 29	65 39 45 95 93	42 58 26 05 27
15 47 44 52 66	95 27 07 99 53	59 36 78 38 48	82 39 61 01 18	33 21 15 94 66
94 55 72 85 73	67 89 75 43 87	54 62 24 44 31	91 19 04 25 92	92 92 74 59 73
42 48 11 62 13	97 34 40 87 21	16 86 84 87 67	03 07 11 20 59	25 70 14 66 70
23 52 37 83 17	73 20 88 98 37	68 93 59 14 16	26 25 22 96 63	05 52 28 25 62
04 49 35 24 94	75 24 63 38 24	45 86 25 10 25	61 96 27 93 35	65 33 71 24 72
00 54 99 76 54	64 05 18 81 59	96 11 96 38 96	54 69 28 23 91	23 28 72 95 29
35 96 31 53 07	26 89 80 93 54	33 35 13 54 62	77 97 45 00 24	90 10 33 93 33
59 80 80 83 91	45 42 72 68 42	83 60 94 97 00	13 02 12 48 92	78 56 52 01 06
46 05 88 52 36	01 39 09 22 86	77 28 14 40 77	93 91 08 36 47	70 61 74 29 41
33 17 90 05 97	87 37 92 52 41	05 56 70 70 07	86 74 31 71 57	85 39 41 18 38
69 23 46 14 06	20 11 74 52 04	15 95 66 00 00	18 74 39 24 23	97 11 89 63 38
19 56 54 14 30	01 75 87 53 79	40 41 92 15 85	66 67 43 68 06	84 96 28 52 07
45 15 51 49 38	19 47 60 72 46	43 66 79 45 43	59 04 79 00 33	20 82 66 95 41
94 86 43 19 94	36 16 81 08 51	34 88 88 15 53	01 54 03 54 56	05 01 45 11 76

Fuente: The Rand Corporation, A Million Random Digits, Free Press, Glencoe, III, 1955, pp. 1–3, con autorización del editor.

Distribución de t

df	Nivel de significación para la prueba de una sola cola					
	.10	.05	.025	.01	.005	.0005
	Nivel de significación para la prueba de dos colas					
	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.265
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Fuente: Es una abreviación del anexo III de Statistical Tables for Biological, Agricultural and Medical Research (ed. 1948), de R.A. Fisher y F. Yates, publicada por Oliver & Boyd, Ltd., Edimburgo y Londres, con la autorización de los autores y editores.

Distribución de X^2

	0 99	0.98	0 95	0.90	0.80	0.70	0.50	0.30	0.20	0 10	0.05	0 02	0.01	0.001
1	0.0'157	0.0'628	0.00393	0.0158	0.0642	0.148	0.455	1.074	1.642	2 706	3.841	5.412	6.635	10.827
2	0.0201	0.0404	0.203	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.815
3	0.115	0.185	0.352	0.584	1.005	1.424	2.366	3.665	4.642	6.251	7.815	9.837	11.341	16.268
4	0.297	0.429	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	13.277	18.465
5	0.554	0.752	1.145	1.610	2.343	3.000	4.351	6.064	7 289	9.236	11.070	13.388	15.086	20.517
6	0.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22 457
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322
8	1.646	2.032	2 167	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15 507	18.168	20.090	26.125
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27 877
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588
11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528
14	4.660	5.368	6.571	7.790	9 467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790
18	7 015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312
19	7633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.620
26	12.198	13.409	15.379	17.292	19 820	21.792	25.336	29 246	31.795	35.563	38.885	42.856	45.642	54.052
27	12.879	14.125	16.151	18.114	20 703	22.719	26.336	30.319	32.912	36.741	40.113	44 140	46.963	55.476
28	13.565	14.847	16.928	18.939	21 588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893
29	14.256	15.574	17.708	19.768	22.475	24.577	28 336	32.461	35.139	39.087	42.557	46.693	49.588	58 302
30	14.953	16.306	18.493	20.599	23.364	25.608	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59 703

Valores de F con significancia de 0.05

N_2 \ N_1	1	2	3	4	5	6	8	12	24	α
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	3.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
α	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

Valores de F con significancia de 0.01

N_2 \ N_1	1	2	3	4	5	6	8	12	24	α
1	40.52	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98.49	99.01	99.17	99.25	99.30	99.33	99.36	99.42	99.46	99.50
3	34.12	30.81	29.46	28.71	28.24	27.91	27.49	27.05	26.60	26.12
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37	13.93	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.27	9.89	9.47	9.02
6	13.74	10.92	9.78	9.15	8.75	8.47	9.10	7.72	7.31	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47	6.07	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67	5.28	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11	4.73	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71	4.33	3.91
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40	4.02	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16	3.78	3.36
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96	3.59	3.16
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80	3.43	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67	3.29	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55	3.18	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45	3.08	2.65
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37	3.00	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30	2.92	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23	2.86	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17	2.80	2.36
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12	2.75	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07	2.70	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03	2.66	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99	2.62	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96	2.58	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93	2.55	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	2.90	2.52	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87	2.49	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84	2.47	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66	2.29	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50	2.12	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.34	1.95	1.38
α	6.64	4.60	3.78	3.32	3.02	2.80	2.51	2.18	1.79	1.00

Serie
APRENDER A INVESTIGAR

ANEXO

Módulo 4

Instructivo para el uso del video

ESTELA URIBE VÉLEZ

1. Uso didáctico del video
2. Videos:
   - *La medición y las ciencias*
   - *La curva normal*
   - *La muestra*

1. USO DIDÁCTICO DEL VIDEO

En este documento queremos mostrar la importancia que tiene la utilización de **video** en las aulas de clase, sus ventajas y desventajas en las actividades académicas. Así mismo, que es responsabilidad del profesor el buen uso que de éste se haga para facilitarle el proceso enseñanza-aprendizaje.

El video es un sistema de almacenamiento de imágenes y sonidos que utiliza los mismos fundamentos técnicos que la televisión y que nació para cubrir las necesidades que las programadoras de T.V. tenían para almacenar sus programas y librarse de la esclavitud de la emisión en directo. Su aparición revolucionó la tecnología televisiva, que hasta el momento se reducía a la transmisión de sucesos en directo, de forma irrepetible. El video configuró una nueva televisión donde el programa elaborado por montaje, el diferido, las repeticiones y los intercambios de programas, enriquecen la producción televisiva.

La mayor ventaja que tiene el video con respecto al resto de los audiovisuales, desde el punto de vista didáctico, es la posibilidad de una **presentación flexible** y un **fed-back inmediato.**

Una de las funciones que debe cumplir el video en el aula es la de ser un facilitador para el proceso enseñanza-aprendizaje.

El docente debe reflexionar sobre la realidad educativa concreta y, como consecuencia de ello, descubrir cuáles son sus necesidades reales con relación al **video** como medio concreto, fruto del desarrollo de las nuevas tecnologías de la comunicación y su aplicación en la enseñanza.

El video como medio didáctico

La forma más frecuente de esta utilización consiste en el trabajo con el grupo/aula –la situación de enseñanza masiva–, en la que el monitor sustituye el discurso del profesor. Generalmente se recurre a este medio para motivar/introducir un tema, aunque para algunos temas y/o disciplinas –generalmente aquellas centradas en la explicación de procesos, etc.– el video introduce algunos momentos didácticos.

El video puede cumplir diferentes funciones dentro de los procesos de enseñanza-aprendizaje; entre ellas tenemos:

Función motivadora

El video, como todos los medios basados en lenguajes visuales, es particularmente apto para transmitir emociones, sensaciones, afectos, que a menudo las palabras no logran expresar con la misma precisión, ya que la imagen, por su misma naturaleza, comunica de manera inmediata, más rápida y

emotiva que la palabra. En este sentido, además de implicar al alumno en la información videográfica, pueden desarrollarse y afirmarse actitudes, estimular la imaginación, la fantasía...

Básicamente se trata de utilizar el video para captar la atención y, a su vez, dar una primera idea, muy general sobre el tema. El tipo de programas que podemos emplear puede proceder de emisiones de televisión comercial, de películas, de actividades realizadas en el propio centro, entre otras. Su grado de elaboración desde el punto de vista educativo puede ser muy elemental, ya que lo fundamental de esto reside en lo que tiene de actualidad o de interés para el alumno. El profesor es quien puede decidir qué tipo de contenidos son los adecuados, en cada caso concreto, para el desarrollo de esta función. El material puede tener los orígenes más diversos.

Función presentadora/introductoria

La presentación del contenido del tema, bien en su totalidad, bien en una parte del mismo, puede ser otra de las funciones del video en el aula. Esta utilización es sensiblemente diferente a la anterior y no se puede entender ni tiene sentido de forma aislada.

El desarrollar un tema en el aula por medio del video debe tener dos acciones previas:

1. Adecuar el programa a la situación concreta de clase.

2. Desarrollar los materiales que van a permitir la comprensión y el desarrollo del mismo, el video por sí solo tiene problemas para, de una forma adecuada, dar una información a un grupo de alumnos. La velocidad de la narración puede ser, entre otros, uno de los problemas que se pueden presentar. Para suplir algunos de estos problemas se hace indispensable que el profesor prepare por lo menos dos tipos de materiales complementarios a esta presentación como son:

   – Materiales de ampliación del contenido del video. Estos deben permitir que el alumno vea la importancia del tema que el video ha mostrado y cómo desarrollar aquellos aspectos tratados en el video que puedan tener un interés relevante. Deben ser fundamentalmente documentos, referencias bibliográficas, etc., que el profesor debe preparar para cada video.

   – Materiales de observación o evaluación del video. Deben ser como guías de observación del problema, en el que se destacan aquellos aspectos del mismo que tienen un mayor interés y que deben ser observados con más atención. Tienen una doble función: destacar los puntos fundamentales y evaluar los contenidos por el alumno. Estas guías contienen una

serie de cuestiones que se han de responder con la información que facilita el programa visionado. No se trata exclusivamente de preguntas y respuestas en el sentido más tradicional. Pueden ser esquemas para completar, mapas a complementar, gráficas, etc.

Un uso complementario de todo lo anterior va a permitir al estudiante no sólo la utilización de cualquier guía, sino también la manipulación de los programas de video utilizados en clase, de tal forma que pueda crear con ellos su propio programa, bien con intención de completar el aprendizaje o bien con una intención evaluadora.

Función informativa

Está directamente relacionada con la adquisición de conocimientos y con la relación que se establece entre las nuevas informaciones que se reciben y las ideas que ya se poseen, desarrollando nuevos conceptos y conocimientos. El video actúa de mediatizador desde el momento que se trata de una observación indirecta. Esta observación no puede limitar el estudio a lo que el entorno le ofrece visualmente, sino que se ocupará de cosas que no pueden ser observadas directamente, bien por problemas especiales –no se puede acceder–, bien por problemas temporales –acontecimientos que pueden apreciarse con la concentración o dilatación del tiempo–, lugares inaccesibles para la visión humana, o por tratarse de una conexión con otras tecnologías –telescopios, microscopios–, o de acceso difícil, costoso o peligroso.

Función instructiva

El video, además de motivar y transmitir información, ha de servir para proporcionar instrumentos tendientes a la organización del conocimiento y al desarrollo de destrezas. Las destrezas, las actitudes de base conseguidas, pueden transferirse a otros ámbitos del conocimiento, de la cultura o de las situaciones vitales a través de principios.

La aplicación de estos principios a la utilización didáctica se centra, fundamentalmente, en la intervención del profesor en la transmisión del mensaje –especialmente, en el modo de presentación–, de tal forma que lo haga más dinámico y activo. Esta intervención requiere un conocimiento de las implicaciones que el modo de presentación de un material tiene en el alumno, la posibilidad de flexibilizar la utilización de los materiales a fin de adecuarlos al mayor abanico de necesidades y condiciones y, por último, requiere también la integración del medio en el contexto.

Estos principios son:

a. El modo de presentación: La eficacia del mensaje depende tanto del contenido como de su presentación.

b. Flexibilidad de utilización: La flexibilidad supone para el profesor el tratamiento de dicho mensaje desde enfoques diversos:

   a. La audiencia: Debe conocer muy bien a su audiencia y tener claras sus necesidades.

   b. Contemplar la necesidad de utilización en situaciones didácticas que no sean solamente grupales.

   c. Utilización integrada en el contexto educativo: El profesor debe presentar contenidos que se integren en el medio afectivo, social y cultural del alumno destinatario.

Función de recapitulador

Este es otro posible uso del video en la situación interactiva de clase, el empleo de las secuencias más significativas del tema expuesto, dándole un tratamiento diferente, más ágil, más breve, puede permitir que todo el tema se resuma en unos pocos minutos, de manera que quede claro cuáles son las ideas y los conceptos fundamentales del mismo.

Las funciones anteriores van interrelacionadas y le corresponde al profesor desarrollarlas de tal forma que se adecuen a sus propósitos; por lo tanto el profesor debe, con anterioridad a la utilización del video en el aula, visionar repetidamente el mismo, analizarlo y decidir cuáles son los puntos esenciales del tema, desarrollando seguidamente los materiales complementarios para los estudiantes.

El video como medio de expresión y de comunicación

Si concebimos el video como un medio que une la comunicación didáctica, es obvio que, además de transmitir información externa –más o menos modificada o manipulada por el profesor–, debe servir de medio de expresión de las propias ideas y experiencias para los protagonistas del proceso enseñanza-aprendizaje –profesor y alumnos–.

El video, por su propia naturaleza, resulta un medio apropiado en una comunicación bidireccional –multidireccional– en el aula. Su desarrollo como tal, exige, no obstante, un cambio radical en algunas concepciones ancladas en el sistema educativo y, especialmente, aquellas relacionadas con la facultad y la libertad para comunicar.

Es necesaria la participación libre y consciente por parte del alumno en el proceso comunicativo.

La simplicidad técnica de manejo y los bajos costos de este medio, hacen posible lo que con el cine era más complicado. La enseñanza de la imagen es hoy factible en su doble vertiente, para contemplarla y para usarla como medio de expresión. Conocer los criterios y principios que están detrás de un mensaje verboicónico nos permite dos tipos de acciones:

– Primero. Facilitar la decodificación completa de los programas que nos llegan por los medios de comunicación de masas, fundamentalmente la televisión.

– Segundo. Permite utilizar un nuevo medio de expresión de forma correcta, de manera que podamos decir lo que deseamos y que nuestro mensaje sea decodificado de acuerdo con nuestra intención.

En relación con esto último debemos recordar que el video es un medio de comunicación, y por lo tanto reúne las condiciones necesarias para poder establecer por medio de él un proceso de comunicación fundamentalmente grupal y en menor medida personal.

El video como medio de investigación

El video reúne unas condiciones que le hacen ser un buen auxiliar para la investigación, tanto para los estudiosos de la educación, profesores, investigadores, etc., como para los propios alumnos. Por lo que respecta a los primeros, el video permite la grabación de distintos tipos de situaciones que tengan que ver con la enseñanza y que, posteriormente, se pueda estudiar al ritmo que se desee. Reuniones de grupos de trabajo, situaciones de clase, entrevistas, etc., son algunas de las ocasiones en las que el video puede aportar la posibilidad de observar aspectos que, de otra manera, pueden pasar inadvertidos.

En el caso de los alumnos que también es extensible a cualquier tipo de investigaciones, el video puede proporcionar el soporte ideal para la observación del desarrollo de procesos de tipo físico, etc., archivo documental, realización de informes y un buen número más de posibilidades que prácticamente está limitado solamente por la imaginación y creatividad de los usuarios.

Otros usos del video

Evidentemente, existen otros usos del video en el aula, entre los cuales podemos destacar su utilización como medio de entretenimiento, o como recursos para la formación permanente del profesorado, para la comunicación universidad-padres, como medio de investigación, entre otros.

Desde este punto de vista, los usos que resultan básicos para los profesores pueden ser:

a. En la formación del profesorado, actuando como instrumento de análisis en diversas técnicas, como microenseñanza, y también en la familiarización de los profesores respecto a dicho medio.

b. En la investigación didáctica, como instrumento de análisis de las interacciones profesor-alumno, de las conductas de los alumnos, de la actuación de los profesores... También constituye un campo de investigación en este terreno la investigación sobre diseño y elaboración de materiales didácticos en video.

Todos los posibles usos descritos casi nunca podremos aislarlos totalmente en la práctica educativa; así por ejemplo, la formación del profesorado no puede desligarse del componente de medio de investigación que el video desarrolla.

Ventajas del video como ayuda didáctica

- Permite mostrar situaciones históricas, presentes y futuras.
- Se pueden repetir acciones.
- Se pueden integrar imagen, color y sonido.
- Permite adecuar parte de un tema en imágenes que se puedan proyectar.
- Mantiene la atención del estudiante, si el tema es motivador.
- Posibilita la reflexión en grupo sobre el tema proyectado.
- Se pueden realizar análisis parciales sobre el tema, suspendiendo la proyección cada vez que el profesor o el estudiante lo consideren necesario.
- Permite la interactividad en clase.
- Puede ser visto en grupos o individualmente.
- No requiere oscuridad para su presentación.

Desventajas del video como ayuda didáctica

- Sólo se debe presentar una idea del tema que se va a tratar.
- El desarrollo del tema puede ser muy rápido y no se capta fácilmente.
- Si el tema es largo y monótono permite que los estudiantes se distraigan.
- Propicia los comentarios durante la proyección, lo que puede distraer a los asistentes.
- Si el tema no está bien tratado permite el adormilamiento entre los asistentes.

Errores más comunes cuando se presenta un video

– No conocer el tema del video con anterioridad a la presentación.
– Desconocer su estado físico (imagen rayada, mal sonido, baches...).
– Pretender que un video por sí solo sea didáctico.
– No preguntar a su audiencia si lo ha visto.
– No hacerle una introducción antes de su presentación.
– No analizar al final de la presentación una evaluación o conclusiones sobre el tema.
– Presentar videos muy largos sin hacer pausas para reflexionar sobre el tema o hacer declaraciones sobre el mismo.
– No adecuar el aula para su presentación.
– Apagar la luz cuando se utiliza un monitor de T.V.
– Utilizar un monitor de T.V. pequeño para una audiencia muy grande.
– Desconocer el funcionamiento de los equipos: T.V., VHS o Betamax.

Algunas recomendaciones para utilizar el video con fines didácticos

QUÉ NO HACER	QUÉ HACER
1. Seleccionar todo el video sólo por el título.	1. Ver el contenido del video y analizarlo para ver si es apropiado al tema que queremos complementar.
2. Presentar un video sin conocer su contenido y su estado físico, sonido, imagen.	2. Revisar muy bien el material seleccionado antes de presentarlo.
3. Proyectar un video sin hacer antes una presentación del tema.	3. Hacer una presentación del tema y explicar qué se pretende con el video.
4. Presentar un video sin preguntar a los participantes si ya conocen su contenido.	4. Debemos interrogar a los participantes si ya conocen el video, para no repetirlo o hacer otro enfoque sobre el mismo.
5. Presentar videos muy largos sin hacer una pausa.	5. Hacer una pausa para evitar la fatiga, hacer preguntas o aclaraciones sobre el tema si es del caso.

6. Presentar videos en idiomas extranjeros,

6. Debemos preguntar si el idioma del video es comprensible para todos.

7. Seleccionar el video como único medio de apoyo a un curso.

7. Debemos combinar medios en nuestros cursos.

8. Utilizar un monitor de T.V. pequeño para audiencias numerosas.

8. Seleccionar el tamaño del monitor de T.V. de acuerdo con el número de la audiencia.

9. Oscurecer el aula cuando se está utilizando un monitor de T.V.

9. Se debe dejar la luz prendida, esto permite observar la reacción de la audiencia ante el video.

10. Terminar la presentación sin hacer conclusiones o preguntas sobre el video.

10. Al terminar de ver un video se deben hacer preguntas, aclaraciones, verificar si se cumplió el objetivo propuesto.

Bibliografía

MARTÍNEZ SÁNCHEZ, Francisco. *La educación ante las nuevas tecnologías de la Educación: Configuración de los videos didácticos*. Anales de la Pedagogía Nº 8, 1990, págs. 159-180.

COLOM CAÑELLAS, Antonio J., SUREDA NEGRE, Jaume, SALINAS IBÁÑEZ, Jesús. *Tecnologías y medios didácticos*. De Cincel S.A., Madrid, 1988.

2. VIDEOS

2.1 VIDEO: «LA MEDICIÓN Y LAS CIENCIAS»

Introducción

El propósito de este video es **mostrar qué es la medición, sus diferentes procesos para recolectar la información.**

El material audiovisual cumple la función de ***complemento o refuerzo*** al material escrito, está concebido como un material autoinstructivo elaborado como imágenes sencillas y de una forma coloquial.

La información consignada en cada uno de los videos es *autosuficiente*, es decir, que en forma independiente comunica una información *completa* sobre un determinado tema. Sin embargo, no debe perder de vista que cada uno de ellos es parte integrante de una unidad global que es todo el curso.

Para complementar el **Módulo 4, Análisis de la información**, se ha elaborado el video **«La medición y las ciencias»**, el cual le dará una visión panorámica y al mismo tiempo será un complemento al tema que usted ha estudiado.

Recomendaciones

Antes de ver este video le recomendamos haber estudiado el Módulo 4, *Análisis de la información*, y estar familiarizado con las técnicas básicas para recolectar la información,

Recuerde que el ideo es una ayuda complementaria, que pretende reforzar el contenido que usted ya estudio, además debe tener en cuenta:

- Para ver el video utilice un monitor de televisión adecuado, mínimo de 14''.
- No oscurezca la sala en donde vea el video, así no se fatigará.
- Recuerde llevar papel y lápiz para que tome nota.
- Es conveniente que sepa manipular el control remoto del VHS o Betamax para que pueda adelantar o retroceder en las ideas que no le sean claras.
- Después de ver el video debe realizar la Autoevaluación que aparece a continuación del guión de contenido. Si ve el video con otros compañeros podrá realizar después una mesa redonda para discutir las respuestas, lo que lo hará más interesante y así los aportes que hagan los participantes serán valiosos.

Guión de contenido del video
«LA MEDICIÓN Y LAS CIENCIAS»

Los investigadores disponen hoy de instrumentos muy precisos para medir magnitudes de espacio, de tiempo, de intensidad, no sólo en nuestro mundo cercano sino en el espacio sideral.

Estos adelantos técnicos son el resultado de un proceso tan largo como la presencia humana sobre la Tierra, tan antiguo como su necesidad de enumerar, de contar, de medir las cosas, los animales, los seres diversos que compartían con él su pequeño mundo circundante. Para numerar cosas o para medir longitudes, lo más obvio era recurrir al propio cuerpo, utilizar los dedos, los brazos, los pies, como unidades de medida; pero al tratarse de longitudes, los pies de Pedro eran más largos que los pies de Pablo, y unos y otros más largos que los de sus mujeres. Fue preciso entonces escoger arbitrariamente el pie derecho de Pedro como medida universal para toda la tribu, y como Pedro no podía concurrir a todos los sitios donde se necesita, se hicieron réplicas de los pies de Pedro para que todos pudieran utilizarlas.

La fábula anterior puede ser o no ser una reseña exacta de la manera como se inició la historia de las medidas en el mundo, pero fue ésta la manera como se impuso, como consecuencia de la revolución francesa, el sistema métrico decimal para las mediciones en casi todos los rincones del planeta. Aunque el metro de cien centímetros tiene relación matemática como una medida de la tierra misma, la medida metro es un objeto de platino e iridio que se encuentra en el museo de «Sévres» cerca de París, esa medida tiene la ventaja de que se le puede tomar réplicas bastante adecuadas, como ésta que es el patrón de longitud para Colombia y reposa en la Universidad Nacional.

Son medidas arbitrarias en su origen, pero aceptadas por el mundo y por consiguiente convertibles para la mayoría de la gente que sabe lo que significa un metro, un kilogramo, una hora, un año luz. La investigación científica moderna se hace en una u otra forma, con base en procesos de medición que pretenden señalar las características de los objetos que se investigan; aunque es preciso decir que lo cuantitativo no constituye la parte más importante de la investigación y que incluso en investigaciones relativas a ciencias sociales la medición tiene sus limitaciones que deben ser comprendidas por el investigador, para que no aspire a llegar con sus conclusiones más allá de lo que le permiten los datos. Los procesos de medición se han desarrollado más allá en las ciencias naturales y en las llamadas exactas que en las ciencias sociales; pero de todos modos siempre y cuando el investigador tome las medidas reservas también el comportamiento de los seres humanos y sus actitudes pueden ser medidas, tanto en las ciencias sociales como en las naturales o en las exactas. Las mediciones que logramos hacer dejan siempre margen al error, porque siempre nos valemos de nuestros sentidos que

están sujetos permanentemente a insuficiencias habituales o temporales, y nuestros instrumentos de medida más perfectos siempre dejan la posibilidad abierta para la incertidumbre. Sin embargo, cuando se toman todas las precauciones y se realizan todas las pruebas y contra pruebas necesarias, el científico puede confiar en sus mediciones y quienes recibimos el beneficio de sus investigaciones también podemos aceptar como sólidas las bases sobre las que están edificadas. Podemos por ejemplo crear hipótesis sobre comportamiento del mercado de algún producto y establecer si dicho comportamiento puede ser observado en sus consecuencias, si luego nuestros sentidos y nuestros instrumentos nos dicen que estamos experimentando lo que preveíamos, podemos tener confianza en que nuestra hipótesis era cierta. Lo que sí tiene que considerar indispensable el investigador para que el proceso de medición tenga su efecto es la clarificación y clasificación de los conceptos que utiliza, de tal manera que haya definiciones que permitan lograr acuerdos acerca de lo que se trata de expresar, o sea, que los otros entiendan lo que quiere decir aunque no lo compartan.

Hay diversos niveles de medición ya sea en física o en astronomía, en sociología o economía, según las propiedades o características que están siendo medidas; ante todo el nivel nominal que sólo sirve para clasificar sin establecer ninguna otra diferencia entre las unidades o individuos de que se trata, y los números vienen a desempeñar medianamente el papel de nombres para distinguir a uno del otro. El nivel ordinal le asigna valores numéricos a los sujetos en tal forma que los valores más altos corresponden a los individuos que tiene más de la característica que se está midiendo, con lo que se diferencia uno del otro, como en el nivel anterior, pero además introduce la posibilidad de un orden determinado. Sin embargo, hay por lo menos dos grandes limitaciones: la diferencia entre el individuo que tiene el número uno y el que tiene el número dos no es igual a la que existe entre éste y el que tiene el número tres; además, no existe el cero absoluto y si le ponemos el número cero a alguno de los muñecos, sería para significar que tiene menor altura que el número uno, pero no significaría ausencia de altura. De esas dos fallas, el nivel intervalo de medición soluciona la primera, permite establecer el orden con intervalos iguales de tal manera que entre 10 y 20 grados hay la misma diferencia que entre 20 y 30 grados de temperatura; pero tampoco tenemos el cero absoluto, pues una temperatura de cero grados no significa carencia de temperatura sino el comienzo de la escala positiva y de la negativa.

Finalmente existe el nivel de razón o radial en la medición, que incluye las posibilidades de los otros niveles, pero también incluye la del cero absoluto. El ingreso de las diversas personas se mide en pesos que son iguales para uno y para otro; por lo tanto, se pueden diferenciar según reciban más o menos. Pero, además, tienen la posibilidad de un cero absoluto como en el caso de quien está desempleado y no recibe un solo centavo de ingreso. El investigador deberá definir muy bien el nivel de medición que corresponde al problema

que desea investigar y el nivel que escoja le estará señalando la técnica estadística que debe utilizar para analizar los fenómenos.

Además de la escogencia adecuada de los niveles de medición que el investigador necesita, en cada caso debe tener bien claros los factores que tomará como constantes y los que tendrá como variables. Generalmente se designan los factores constantes con las primeras letras del alfabeto castellano o del griego, y las variables con las últimas letras de nuestro alfabeto, a las que se puede añadir distintivos numéricos. Por último, las variables pueden tomar sólo cierto número de valores y entonces son variables discretas, como cuando damos un valor unitario a cada respuesta acertada de un alumno, o pueden ser variables que tomen un número infinito de valores y entonces son variables continuas como en el sistema métrico decimal.

Volvamos sobre la posibilidad de error que siempre acecha al investigador y que es más preocupante según sea la necesidad de precisión y el instrumento que utiliza. Los errores pueden obedecer a falla personal, como por ejemplo: un defecto de la vista en el investigador, o a defectos sistemáticos surgidos de fallas intencionadas o fortuitas de los aparatos de medición. Y finalmente se dan los errores aleatorios, que corresponden al mismo proceso de medición, ya que por ejemplo: la más precisa de las balanzas no nos da siempre la misma medida, aunque es imposible cerrar el paso a toda posibilidad de error. El científico tiene que cuidar que su investigación sea siempre confiable, válida y representativa.

La calidad interna radica en la representatividad o capacidad de ser generalizada que tenga la investigación en los resultados que exhibe. La confiabilidad de una medición es la consistencia que muestra al ser realizada varias veces, o sea se refiere a la ausencia de errores que aparece en pruebas diferentes.

La validez de una medición es el grado con que ésta representa una medida de lo que realmente se quiere medir, o sea de aquello que en etapa anterior se ha aclarado conceptualmente como objeto de la investigación o de la medición específica; con lo que ya se está diciendo que es más importante asegurarse de la validez de la medición que de su confiabilidad. Aunque no es posible establecer directamente la validez de una medición, la práctica científica ha consagrado varias maneras indirectas, ya sea para mediciones hechas en el campo de las ciencias naturales o exactas, o ya sea en el de las ciencias sociales.

Por último la representatividad o generalidad de una medición tiene que ver con el grado en que sus resultados surgidos a partir de una muestra pueden ser atribuidos a la población general; porque si la medición no está bien hecha pierde representatividad así se haya escogido cuidadosamente la muestra.

Autoevaluación

Apoyándose en el contenido del Módulo 4 y del video que acaba de ver, conteste o complete si es del caso las siguientes preguntas:

1. ¿Qué debe considerar el investigador para que el proceso de medición tenga efecto?

2. ¿Cómo pueden ser los niveles de medición?

3. ¿Para qué sirve el nivel nominal?

4. ¿Para qué sirve el nivel ordinal?

5. ¿En qué consiste el nivel de razón o radial en la medición?

6. ¿Qué determina la técnica estadística que el investigador debe utilizar para analizar los fenómenos?

7. ¿Qué otros factores debe considerar el investigador además de la escogencia de los niveles de medición?

8. ¿Cómo pueden ser las variables?

9. ¿Qué diferencia hay entre estas variables?

10. ¿A qué puede obedecer la posibilidad de error en la medición que realiza el investigador?

11. ¿De qué depende la calidad interna de un proceso de investigación?

12. ¿En qué radica la calidad externa de los procesos de investigación?

13. ¿En qué consiste la confiabilidad de una medición?

14. ¿Qué es la validez de una medición?

Respuestas a la Autoevaluación

1. Clarificación y clasificación de los conceptos que utiliza, para lograr acuerdos acerca de lo que se trata de expresar.
2. Nominal, ordinal y el nivel de razón o radial.
3. Para clasificar sin establecer ninguna otra diferencia entre las unidades o individuos de que se trate.
4. Asigna valores numéricos y puede introducir la posibilidad de un orden determinado.
5. Incluye las posibilidades de los otros niveles y la del cero absoluto.
6. El nivel de medición.
7. Los factores que tomará como constantes y los que tendrá como variables.
8. Discretas y continuas.
9. Las discretas tienen número de valores, las continuas tienen un número infinito de valor.
10. • A falla personal.
    • Defectos sistemáticos surgidos de fallas intencionadas o fortuitas de los aparatos de medición.
    • Errores aleatorios que corresponden al mismo proceso de medición.
11. De la confiabilidad y validez que tenga la observación de los datos.
12. En la representabilidad o capacidad de ser generalizada que tenga la investigación en los resultados que exhibe.
13. Es la consistencia que muestra al ser realizada varias veces, o sea a la ausencia de error que aparece en pruebas diferentes.
14. Es el grado con que ésta representa una medida de lo que realmente se quiere medir.

Después de haber realizado la autoevaluación puede comparar sus respuestas con el guión de contenido.

2.2 VIDEO: «LA CURVA NORMAL»

Introducción

El propósito de este video es **mostrar qué es la curva normal, su comprensión y la aplicación en el trabajo investigativo.**

El material audiovisual cumple la función de ***complemento o refuerzo*** al material escrito, está concebido como un material autoinstructivo elaborado con imágenes sencillas y de una forma coloquial.

La información consignada en cada uno de los videos es *autosuficiente,* es decir, que en forma independiente comunica una información *completa* sobre un determinado tema. Sin embargo, no debe perder de vista que cada uno de ellos es parte integrante de una unidad global que es todo el curso.

Para complementar el **Módulo 4. Análisis de la información,** se ha elaborado el video **«La curva normal»,** el cual le dará una visión panorámica y al mismo tiempo será un complemento al tema que usted ha estudiado.

Recomendaciones

Antes de ver este video que tiene una duración de 15 minutos, le recomendamos haber estudiado el Módulo 4, conocer en qué consiste el modelo teórico de la curva normal y cuáles son sus múltiples aplicaciones.

Recuerde que el video es una ayuda complementaria, que pretende reforzar el contenido que usted ya estudió; además tenga en cuenta:

- Para ver el video utilice un monitor de televisión adecuado, mínimo de 14”.
- No oscurezca la sala en donde vea el video, así no se fatigará.
- Recuerde llevar papel y lápiz para que tome nota.
- Es conveniente que sepa manejar el control remoto del VHS o del Betamax para que pueda adelantar o retroceder en las ideas que no le sean claras.

Después de ver el video debe realizar la **Autoevaluación** que aparece a continuación del guión de contenido. Si ve el video con otros compañeros podrá realizar después una mesa redonda para discutir las respuestas, lo que lo hará más interesante y así los aportes que hagan los participantes serán valiosos.

Guión de contenido del video
«LA CURVA NORMAL»

¿Qué nuevas vías son necesarias para la ciudad? ¿Qué cantidad de provisiones debe almacenar este negocio para atender a sus clientes? ¿En qué nivel está la calidad de esta deportista? Para responder preguntas como éstas es necesario observar y registrar la frecuencia y la forma como ocurren ciertos hechos a lo largo de un período. En el caso de las vías, por ejemplo: podríamos llegar a conocer datos como el promedio de vehículos que transitan durante varios días de la semana; o la clase de vehículos que con más frecuencia pasan por allí; son buses, camiones o carros particulares, y con mayor observación podríamos llegar a conocer muchos detalles de la forma como se comporta el tránsito en esa calle en distintas horas y en diversas épocas del año. De la misma manera y con procesos similares de observación el propietario de este negocio podría calcular las provisiones que debe almacenar de acuerdo con la forma como se comporta la demanda de sus clientes según días, horas y clase de producto.

En el caso del deportista, también sería posible registrar las marcas que logra en su deporte, con el fin de medir sus progresos a lo largo de un entrenamiento o para compararlo con las marcas que logran otros competidores. La estadística nos proporciona entonces un conjunto de técnicas para registrar y observar el comportamiento de fenómenos variables. Cuando se efectúan mediciones estadísticas se observan diferentes formas en la frecuencia con que ocurre el fenómeno. Por ejemplo: si examinas el comportamiento del cliente en esta panadería a lo largo del día, es probable encontrar que las ventas máximas se localizan en dos períodos, en las primeras horas de la mañana y las últimas de la tarde. Esta clase de distribuciones se conocen como *bimodales.* En otro tipo de distribuciones las mayores frecuencias se acumulan a uno u otro lado de los valores de la variable. Por ejemplo: es probable que las calificaciones que un profesor oneroso concede a sus alumnos estén representadas por un tipo de curva concreta en la cual se registra que la mayor parte de estudiantes consiguen buenas notas; en el caso del profesor muy estricto la distribución de las frecuencias será opuesta al caso anterior.

En el caso dos, la curva refleja que sólo unos pocos estudiantes consiguen buenas notas, mientras que la mayoría sólo alcanzan calificaciones malas o mediocres. Otra clase de distribución, conocida como *distribución normal,* es aquella en la cual las frecuencias tienden a concentrarse uniformemente alrededor de un valor central. A esa distribución se aproxima por ejemplo: los puntajes que los estudiantes consiguen en exámenes como las pruebas del ICFES, o los de ingreso a la universidad; en ese caso la mayor parte de estudiantes se concentra en los puntajes intermedios de la escala, mientras que un grupo menor consigue los puntajes máximos y otros resultantes se ubican en los mínimos.

A partir del concepto de distribución normal, la ciencia de la estadística ha desarrollado un modelo teórico de importantes y múltiples aplicaciones. Este modelo se conoce como *curva normal* y no es otra cosa que una representación gráfica y matemática de aquellos fenómenos que bajo ciertas condiciones pueden considerarse como normalmente distribuidos. En su origen la curva normal surge como representación de la frecuencia de eventos que ocurren al azar, y por lo tanto como parte de la teoría de la probabilidad. En efecto, al tomar estas ocho monedas podemos preguntarnos con qué frecuencia relativa saldrán caras si las lanzáramos un gran número de veces. Lo primero que responde la teoría de probabilidad es que en la mayor parte de los casos saldrán cuatro caras, del mismo modo podremos suponer que en el menor número de veces saldrán solamente caras y que en un número de veces igualmente menor no saldrán caras. El histograma completo de frecuencias relativas para el caso propuesto tendrá finalmente esta distribución, y si en vez de ocho lanzáramos dieciséis monedas aumentaría el número de las barras dentro del histograma y cada vez se identificaría mejor el perfil de la curva y de la distribución normal. De esta manera la curva normal representa en su origen la frecuencia con que ocurren fenómenos regidos por el azar, sobre los cuales operan multitud de factores independientes. De la misma forma que podemos encontrar círculos de diferente diámetro sin que por ello pierdan su carácter de círculos, es posible también encontrar curvas cercanas a la normal, pero con diferentes configuraciones. Por ejemplo: estas dos curvas registran el mismo valor como promedio pero su amplitud varía porque es distinta a su desviación estándar.

Estas dos curvas registran promedios diferentes pero sus desviaciones estándar son iguales.

En estas dos curvas los promedios y las desviaciones estándar son diferentes.

Sin embargo la familia de las curvas normales cumplen las siguientes características: en primer lugar toda curva normal es simétrica, es decir que la mitad de su lado izquierdo es igual a la mitad del lado derecho, igualmente la distribución que representa la curva normal es unimodal a tal punto que promedio, mediana, y moda tienen todos el mismo valor.

De otra parte, y a diferencia de un polígono de frecuencia, la curva normal es continua, y por lo tanto sólo registra en rigor valores de variables que sean continuas.

Otra característica común es que la curva normal es asintótica, es decir, que no toca el eje de las «x» y que por lo tanto ocupará un número infinito de observaciones; si bien es cierto que en las ciencias sociales las mediciones no son continuas ni el número de observaciones es infinito, la utilidad práctica de la curva normal es su valor como modelo estadístico y matemático.

La curva normal es un instrumento que ayuda a organizar, describir y predecir el comportamiento de fenómenos que en su distribución se aproximan a la distribución normal. Así por ejemplo, y tomando la curva normal como modelo, podrían estimarse el porcentaje de colombianos adultos cuya estatura varía entre 1.60 metros y 1.80 metros y el porcentaje de aquellos que miden menos de 1.60 metros o más de 1.80 metros.

Otra forma en que la Curva Normal actúa como modelo de suma importancia es en la llamada *Distribución de Muestreo.* En efecto, si obtenemos una gran cantidad de muestras aleatorias de un universo dado, el valor de los promedios y de otras estadística de esas muestras tienden a conformar una distribución normal. A modo de ejemplo veamos este caso: los valores 1, 2 y 3 conforman nuestro universo. De ese universo vamos a obtener todas las posibles muestras de dos elementos. He aquí la primera muestra, el promedio de esa muestra es uno y lo anotamos en una tabla de frecuencia. Ahora vamos a repetir el procedimiento con las otras posibles muestras, es decir, tomamos los valores de la muestra y consignamos en la tabla el valor de su promedio, he aquí el resultado final, los promedios de las diferentes muestras tienen una distribución muy cerca a la normal. Si el universo fuera mayor y las muestras más grandes el perfil de la curva normal sería más notable. Hay otro hecho en extremo interesante y es que el valor promedio de la distribución de estos promedios coinciden con el valor promedio del universo original. Este carácter normal de la distribución de muestreo es de gran importancia para realizar inferencias ya que permite al investigador calcular la probabilidad de que el valor de una estadística obtenida en una muestra, coincida con el valor del parámetro de la población o universo. Como modelo teórico, la curva normal tiene en síntesis las siguientes características: como sabemos es una distribución simétrica y asintótica que no toca el eje horizontal y se extiende indefinidamente; el área total debajo de la curva puede representar el número total de elementos de la población o universo, es decir, constituye el 100% de los casos; esta propiedad permite que entre dos puntajes a lo largo del eje horizontal sea posible determinar la proporción de la frecuencia entre dos puntajes; proporción que será igual al área de la curva entre esos dos puntos. Cuando se normaliza una distribución, el medio aritmético, la mediana y la moda son iguales a cero; por lo tanto, los valores al lado izquierdo del eje horizontal son negativos mientras que los valores del lado derecho son positivos.

La desviación estándar en una desviación normalizada es igual a uno, el 68.26% de los valores de la distribución se localizan bajo el área determinada por una desviación estándar a cada lado del medio aritmético. Del mismo modo el 95.46% de los valores se localizan entre extremos de dos desviaciones a lado y lado del medio. El 99% de los valores se agrupan bajo el área que determinan tres desviaciones estándar a lado y lado del mismo medio. La comprensión y aplicación de la curva normal es una exigencia en el trabajo investigativo, particularmente cuando se desea hacer inferencias a partir de

una muestra. La curva es un modelo, es una abstracción; para aplicarla con acierto se requiere su exacta comprensión como instrumento estadístico; lo mismo que el conocimiento preciso de la índole y características de la investigación para la cual se aplique.

Autoevaluación

Apoyándose en el contenido del Módulo 4, y en el video que acaba de ver conteste o complete si es del caso las siguientes preguntas:

1. ¿Qué es la distribución normal?
2. ¿Qué es la curva normal?
3. ¿Por qué surge la curva normal?
4. ¿Cuales son las características que deben cumplir la familia de las curvas normales?
5. ¿Qué es una curva normal asintótica?
6. ¿Cuál es la utilidad práctica de la curva normal?
7. Como modelo estadístico y matemático la curva normal es un instrumento que ayuda a...
8. ¿Qué es distribución de muestreo?
9. ¿Qué características tiene la curva normal como modelo teórico?
10. Cuando se normaliza una distribución, la media aritmética, la mediana y la moda son igual a....
11. ¿A qué es igual la sesviación estándar de una distribución normalizada?
12. ¿Qué se necesita para aplicar la curva normal con acierto?

Después de haber contestado las preguntas puedes comparar sus respuestas con el guión de contenido.

Respuestas a la Autoevaluación

1. Es aquella en la cual las frecuencias tienden a concentrarse uniformemente alrededor de un valor central.
2. Es una representación gráfica y matemática de aquellos fenómenos que bajo ciertas condiciones pueden considerarse como normalmente distribuidos.
3. Como representación de la frecuencia de eventos que ocurren al azar y por lo tanto como parte de la teoría de la probabilidad.
4. – Toda curva normal es simétrica.
   – La distribución que representa la curva normal es la unimodal.
   – La curva normal es continua.
   – La curva normal es asintótica.
5. Es la que no toca el eje de la X y por lo tanto ocupará un número infinito de observaciones.
6. Es su valor como modelo estadístico y matemático.
7. Organizar, describir y predecir el comportamiento de fenómenos que en su distribución se aproximan a la distribución normal.
8. Cuando la curva normal actúa como modelo se obtiene una gran cantidad de muestras aleatorias de un universo dado y el valor de los promedios y de otras estadísticas tiende a conformar una distribución de muestreo.
9. – Es una distribución simétrica y asintótica que no toca el eje horizontal y se extiende independientemente.
   – El área total debajo de la curva normal puede representar el número total de elementos de la población urbana.
10. Cero
11. A uno
12. Su exacta comprensión como instrumento estadístico, lo mismo que el conocimiento preciso de la índole y características de la investigación para la cual se aplica.

2.3 VIDEO: «LA MUESTRA»

Introducción

El propósito de este video es **mostrar qué es la muestra, cuándo utilizar una muestra, qué obtiene el investigador a través de la muestra.**

El material audiovisual cumple la función de ***complemento o refuerzo*** al material escrito, está concebido como un material autoinstructivo elaborado con imágenes sencillas y de una forma coloquial.

La información consignada en cada uno de los videos es *autosuficiente*, es decir, que en forma independiente comunica una información *completa* sobre un determinado tema. Sin embargo, no debe perder de vista que cada uno de ellos es parte integrante de una unidad global que es todo el curso.

Para complementar el **Módulo 4. Análisis de la información**, se ha elaborado el video: **«La muestra»**, el cual le dará una visión panorámica y al mismo tiempo será un complemento al tema que usted ha estudiado.

Recomendaciones

Antes de ver este video que tiene una duración de 15 minutos, le recomendamos haber estudiado el Módulo 4, conocer qué es la muestra, cómo se utiliza, para qué sirve, cómo se determina ésta. Recuerde que el video es una ayuda complementaria, que pretende reforzar el contenido que usted ya estudió. Además tenga en cuenta:

- Para ver el video utilice un monitor de televisión adecuado, mínimo de 14".

- No oscurezca la sala en donde vea el video, así no se fatigará.

- Recuerde llevar el papel y lápiz para que tome nota.

- Es conveniente que sepa manejar el control remoto del VHS o del Betamax para que pueda adelantar o retroceder en las ideas que no le sean claras.

Después de ver el video debe realizar la **Autoevaluación** que aparece a continuación del guión de contenido. Si ve el video con otros compañeros podrá realizar después una mesa redonda para discutir las respuestas, lo que lo hará más interesante y así los aportes que hagan los participantes serán valiosos.

Guión de contenido del video
«LA MUESTRA»

Muchas afirmaciones sobre características de las personas de una región o sobre las que tienen una misma profesión u oficio, o sobre características de las personas según su edad o sexo son ejemplo de generalizaciones, hechas a partir de la observación de apenas unos casos. Eso quiere decir que en la vida cotidiana deducimos que una población tiene cierta característica por el hecho de haberla observado en una muestra de esa población.

Todos generalizamos pero todos sabemos que esas generalizaciones no son científicamente ciertas; sin embargo, la ciencia también hace generalizaciones a partir de la observación de un número limitado de casos. La diferencia está en que la ciencia sigue rigurosos métodos para garantizar la validez de sus generalizaciones. En efecto, el trabajo del investigador científico es encontrar leyes, relaciones, comportamientos y otras características que permitan conocer mejor, bien sea el mundo de los fenómenos naturales o bien el mundo de los fenómenos humanos y sociales. En este proceso son muchas y diversas las preguntas que motivan un trabajo de investigación. Por ejemplo: ¿cuántos desempleados tiene el país? o ¿cómo puede controlarse la roya de los cafetos?, o ¿de qué manera influye el nivel educacional de los padres en el desarrollo físico y mental de sus hijos? En casos como estos los desempleados o los cafetos o los padres de familia constituyen la población o universo objeto de cada investigación. Pero ante la imposibilidad de que el investigador examine a cada individuo de su universo o población sean desempleados, o cafetos, o padres de familia o cualquier otro conglomerado, se hace indispensable que tan solo observe una muestra tomada de la población o universo objeto de su estudio y que a partir de esa muestra pueda generalizar válidamente para todos los casos. Así pues, podemos decir que cuando el investigador desea conocer una característica de una población o universo su trabajo se orienta en primer lugar a observar esa característica en una muestra tomada de esa población o universo, los valores que el investigador mide en la muestra se conocen como *estadística* y los valores que corresponden a la población o universo se conocen como *parámetros.* Así por ejemplo puede ocurrir que en una muestra encontremos un 10% de desempleo. Ese valor sería la estadística, pero es posible que el valor real del desempleo en la ciudad en la cual se tomó la muestra al contabilizar absolutamente todos los casos sea un 12%, esta cifra entonces sería el parámetro. Podemos decir entonces que el investigador obtiene estadísticas de la muestra para inferir parámetros de la población o universo. En este proceso de inferencia científica, el investigador debe cumplir varias y rigurosas exigencias; debe en primer lugar garantizar que su muestra sea representativa del universo, es decir, que la muestra reproduce las distribuciones y valores del universo o poblaciones con márgenes de error calculables. La representatividad de la muestra depende de varios factores; uno de ellos es la identificación del universo de don-

de va a sacar la muestra, esta identificación del universo o marco de muestreo no siempre es sencilla; hay algunas dificultades inherentes, la primera es la posibilidad de delimitar un universo, por el hecho de que todo universo es un organismo vivo que se expande, que crece y es dinámico; tomemos el caso de la población de una ciudad o el número de carros que hay en la ciudad, o el número de maestros que hay en una nación, esos universos están permanentemente creciendo y por lo tanto es muy difícil dar una delimitación precisa en un momento preciso; en segundo lugar existe una segunda clase de limitaciones que tiene que ver con el tipo de objeto social sobre el cual se va a hacer la muestra.

Hay objetos sociales que de por sí son distintamente definidos como por ejemplo: el número de alcohólicos que hay en una ciudad o el número de vagabundos, etc., en esos casos evidentemente no hay manera de precisar el universo; hay una tercera clase de delimitaciones que tienen que ver con las posibilidades tecnológicas al servicio de una investigación, para poder delimitar ese universo en objetos que están ya centralizados, que tienen registros públicos, por ejemplo el problema no es tanto el dato ni la clase de dato, sino el control y la sofisticación que existe en cuanto a la información para ese tipo de datos. Otro factor influyente en la representatividad de la muestra es el diseño muestral. Diseñar una muestra es buscar que cada componente de un universo tenga la posibilidad de estar incluido en la muestra y que se pueda calcular la posibilidad que cada elemento tenga de estar en la muestra.

Las muestras aleatorias o probabilísticas son de varias clases y exigen diferentes procedimientos de selección; sin embargo no son las únicas que se utilizan en la investigación porque también existen muestras no probabilísticas. Las muestras no probabilísticas o intencionales se justifican particularmente en etapas exploratorias de la investigación o para finar instrumentos de redacción de datos o para estudios en profundidad.

Este es un caso: la muestra no probabilística se elabora siempre sin tener en cuenta los procedimientos estadísticos de muestreo, es aquella donde no se conoce el universo y es muy difícil, por lo tanto, determinar una muestra, y donde los elementos de la muestra tienen no todas las mismas posibilidades de ser escogidas para el estudio. Vamos a poner un ejemplo muy claro que es la forma de explicarlo. En investigación tenemos una forma de conseguir información llamada bola de nieve. La podemos aplicar a una investigación en pintura. Por ejemplo: queremos investigar a los pintores en Colombia, únicamente conocemos un pintor y no conocemos a los demás, por lo tanto no conocemos ni el universo ni la muestra. Vamos donde este pintor que conocemos y hacemos la entrevista, al final de la entrevista le preguntamos por dos pintores similares a él en el grupo que nos interesa. Para hacerle la entrevista vamos donde esta gente, los entrevistamos y al final les preguntamos también si conocen otros dos pintores, cada uno de ellos. En esta forma va-

mos ampliando el grupo de gente que estamos necesitando y la bola de nieve va creciendo. Ya no tenemos solamente uno, tenemos seis y eso se va ampliando. Cada persona nos va dando dos nombres más y en esta forma vamos creando un universo de donde podemos sacar una muestra que la vamos creando. Inicialmente en este caso sería la muestra primero y el universo después, pero únicamente debe ser utilizada la muestra no probabilística en aquellos casos en que no exista conocimiento sobre el universo y por lo tanto la dificultad para conseguir una muestra sea muy grande.

En resumen el investigador define con la mayor claridad posible el universo de su estudio, que las amas de casa, que los miembros de partidos políticos, que los de un servicio público, etc., ahora el investigador obtiene muestras unas veces aleatorias y otras intencionales de la población o universo que desea estudiar, pero ojo, que sean muestras representativas.

El investigador hace sus observaciones y mediciones sobre la muestra, de esa manera obtiene estadísticas o sea valores como frecuencias, promedios y otros que le sirven para conocer el comportamiento de la muestra de las variables que está investigando. Finalmente y mediante procesos estadísticos el investigador puede, con margen de errores calculables, generalizar a la población o universo las observaciones hechas a la muestra; de modo general eso es así; pero hay unos pasos de análisis de datos y comprobación de hipótesis muy rigurosos. Medir algo en la muestra y generalizar a la población es lo que se conoce como *inferencia,* es decir, que una estadística nos permite hallar un parámetro.

Para definir el tamaño de la muestra hay que considerar algunos aspectos, unos de orden práctico y otros de orden teórico. En el orden práctico y para definir su muestra busca orientaciones en otros estudios similares e igualmente examina los recursos económicos con que cuenta, el tiempo de que dispone. Desde el punto de vista teórico, el investigador determina su muestra principalmente de acuerdo con los objetivos de su estudio e igualmente se sirve de las fórmulas que las estadísticas proporcionan para calcular tamaño de muestra. ¿Cómo influyen estos factores prácticos y teóricos para que un investigador determine el tamaño de su muestra?

Un factor importante es la naturaleza del estudio o sea qué tipo de investigación se está realizando, de qué tipo de investigación se trata. Pongamos algunos ejemplos:

En primer lugar es muy importante la naturaleza del estudio, es decir, qué tipo de investigación se está realizando, en qué tipo de disciplina y cuales son los objetivos precisos de la investigación para el momento en que el investigador selecciona o determina la muestra. En segundo lugar hay dos factores también muy importantes, el factor económico y el factor tiempo que ya el

investigador debe haber tenido en cuenta desde el momento en que inició su investigación. La complejidad de una muestra por ejemplo hace que aumente el precio de una investigación o los costos de una investigación. Por otra parte, también se puede racionalizar adecuadamente el tiempo si se han hecho estudios previos sobre lo que es la muestra. También debe tenerse en cuenta el tipo de procesamiento de la información que se va a hacer, y por otra parte la forma como se va a hacer la recolección de información. Una tercera característica importante es la tradición que otros investigadores han construido con relación al tipo de muestra que han utilizado, para investigaciones o estudios similares, finalmente la estadística y la teoría del muestreo han proporcionado herramientas muy valiosas a los investigadores de diferentes disciplinas para la adecuada toma de decisiones, en el momento de determinación de la muestra.

La precisión de los resultados no sólo depende del tamaño de las muestras, si hay errores de recolección, una muestra grande podría ser mala, mientas que una muestra pequeña con cuidadosa recolección y selección podría ser excelente.

Una muestra grande no es suficiente para garantizar la precisión de los resultados.

Para determinar una muestra no basta con saber fórmulas matemáticas que nos permitan encontrar su tamaño, depende en primer lugar de la precisión con que esté definido el universo de la investigación, de la claridad que el investigador tenga sobre los objetivos y de su estudio, la recolección y procesamiento cuidados de los datos. De esta forma la claridad conceptual y rigor metodológico son la garantía necesaria para que finalmente las técnicas estadísticas permitan inferir resultados a partir de una muestra.

Autoevaluación

Apoyándose en el contenido del Módulo 4, y en el video que acaba de ver conteste o complete si es del caso las siguientes preguntas:

1. ¿Qué es una muestra?
2. ¿En qué se diferencian las generalizaciones que hace la ciencia con respecto a las generalizaciones que no son científicamente ciertas?
3. ¿Cómo se llaman los valores que el investigador mide en la muestra?
4. ¿Con qué nombre se conocen los valores que corresponden a la población o universo?
5. ¿De qué depende la representatividad de la muestra?
6. ¿Qué es diseñar una muestra?
7. ¿Qué tipos de muestras se utilizan en la investigación?
8. ¿Cuándo se utilizan las muestras no probabilísticas o intencionales?

9 ¿Qué es la muestra no probabilística?

10. El investigador hace sus observaciones y mediciones sobre...
11. ¿Qué obtiene el investigador a través de la muestra?
12. ¿,Qué es inferencia?
13. ¿Cómo se define el tamaño de la muestra?
14. Para determinar una muestra ¿qué debe tener en cuenta un investigador?
15. ¿Qué constituye la garantía necesaria para que finalmente las técnicas estadísticas permitan resultados a partir de una muestra?

Respuestas a la Autoevaluación

1. Son las generalizaciones hechas a partir de la observación de apenas unos casos.

2. En que la ciencia sigue rigurosos métodos para garantizar la validez de sus generalizaciones.

3. Estadísticos.

4. Parámetros.

5. De la identificación del universo de donde se va a sacar la muestra.

6. Es buscar que cada componente de un universo tenga la posibilidad de estar incluido en la muestra y que se pueda calcular la posibilidad que cada elemento tenga de estar en la muestra.

7. <->Muestras aleatorias o probabilísticas

   <->Muestras no probabilísticas o intencionales

8. Se justifican particularmente en etapas exploratorias de la investigación o para afinar instrumentos de redacción de datos o para estudios en profundidad.

9. Es aquella donde no se conoce el universo y es muy difícil por lo tanto determina una muestra y donde los elementos de la muestra no tienen todos la misma posibilidad de ser escogidos para el estudio.

10. La muestra

11. Obtiene estadística, o sea valores como frecuencias, promedios y otros que le sirven para conocer el comportamiento de la muestra de las variables que está investigando.

12. Es medir algo en la muestra y generalizar a la población.

13. Se debe buscar orientación en otros estudios similares, examinar los recursos económicos con que se cuenta, el tiempo de que se dispone, lo anterior desde el punto de vista práctico; y desde el punto de vista teórico el investigador determina su muestra principalmente de acuerdo a los objetivos de su estudio e igualmente se sirve de las fórmulas que la estadística proporciona para calcular tamaños de muestra.

14. – La precisión con que esté definido el universo de la investigación.
    – La claridad que el investigador tenga sobre los objetivos y variables de su estudio.
    – La recolección y procesamiento cuidadoso de los datos.

15. La claridad conceptual y rigor metodológico.

Después de haber contestado las preguntas puedes comparar sus respuestas con el guión de contenido.

Bibliografía

BERENSON, Mark. *Estadística básica en administración. Conceptos y aplicaciones*. 1992.

BRIONES, G. *Métodos y técnicas de investigación para las ciencias sociales*. México, Trillas, 1982.

CARVAJAL, Lizardo. *Metodología de la investigación*. Séptima edición. Cali, FAID, 1991.

CEBALLOS ROJAS, Argemiro. *Estadística descriptiva y probabilidad básica*. 1994.

GODINO, J. y otros. *Azar y probabilidad. Fundamentos didácticos y propuestas curriculares*. Síntesis, 1991.

GONZÁLEZ REYNA, Susana. *Manual de redacción e investigación documental*. Segunda edición, México, Trillas, 1994.

HARNETT y MURPHY. *Introducción al análisis estadístico*. Addison-Wesley Iberoamericana, 1993.

HOPKINS, Kenneth. *Estadística básica para las ciencias sociales y el comportamiento*. 1997.

KREYSZIG, Erwin. *Estadística matemática*. Limusa. 1983.

MASON, Robert; LIND Douglas. *Estadística para administración y economía*. 8ª edición. Santa Fe de Bogotá, Alfaomega Grupo Editor S.A. de C.V., 1998.

MENDENHALL, William. *Introducción a la probabilidad y a la estadística*. Grupo Editorial Iberoamericana, 1991.

MENDENHALL, William; WACKERLY, Dennis; SCHEAFFER, Richard. *Estadística matemática con aplicaciones*. México, Grupo Editorial Iberoamericana, 1994.

MENDENHALL, SCHEAFFER y WACKERLY. *Estadística matemática con aplicaciones*. Grupo Editorial Iberoamericana, 1986.

MONTGOMERY, Douglas C. *Diseño y análisis de instrumentos*. Grupos Editorial Iberoamericana, 1991.

MONTGOMERY, Douglas C. *Control estadístico de la calidad*. Grupo Editorial Iberoamericana, 1991.

PARZEN, ENMANUEL. *Teoría moderna de probabilidades y sus aplicaciones*. México, Limusa, 1991.

PEÑA SÁNCHEZ DE RIVERA, Daniel. *Estadística, modelos y métodos*. Tomo I y Tomo II. Alianza Editores, 1990.

SCHEAFFER, MENDENHALL, OTT. *Elementos de muestreo*. Grupo Editorial Iberoamericana, 1987.

SCHEAFFER y McCLAVE. *Probabilidad y estadística para ingeniería*, Grupo Editorial Iberoamericana, 1993.

SUÁREZ DE LA CRUZ, Alberto Camilo. *Metodología para el estudio y la investigación*. Cuarta edición, Bogotá, Ediciones Ciencias y Derecha, 1991.